RAND Journal of Economics
Vol. 31, No. 3, Autumn 2000
pp. 488–501

Two-stage patent races and patent policy

Vincenzo Denicolò*

I analyze the optimal degree of forward patent protection in a two-stage patent race framework. I compare three patent regimes, as the second innovation may be unpatentable and infringing (UI), patentable and infringing (PI), or patentable and not infringing (PN). Forward protection is highest in regime UI and lowest in regime PN. I identify a fundamental inefficiency affecting regime UI, namely that it always leads to underinvestment in the second innovation, and I note various determinants of the welfare ranking of the regimes. Specifically, strong forward protection becomes less attractive as the relative profitability of the first innovation increases and the relative difficulty of obtaining it decreases.

1. Introduction

■ The debate on optimal patent design has recently focused on the cumulative nature of innovation, which means not only that innovators must be protected from imitators (backward protection) but also that they should be rewarded for opening the way to subsequent technical improvements (forward protection). In an influential article, Green and Scotchmer (1995) argue that first-generation innovators should be given strong forward protection so as to overcome the intertemporal externality that arises when second-generation improvements can be obtained by outsiders. Similar concerns are voiced by Chang (1995), Scotchmer (1996), and Matutes, Regibeau, and Rockett (1996). However, broad forward protection can stifle second-generation products, as is emphasized by Merges and Nelson (1990, 1994) and Heller and Eisenberg (1998).

This article revisits the two-stage research problem but departs from the previous literature by assuming that there is a patent race at each stage. There are two main differences between my assumption and Green and Scotchmer's hypothesis that one firm has the basic idea and another firm is uniquely capable of obtaining the improvement. First, I allow for competition in R&D. (Green and Scotchmer also assume that first- and second-generation innovators can reach licensing agreements before the second-generation firm invests. If several firms are racing to achieve the second innovation, this assumption becomes problematic, and so I rule out *ex ante* agreements.) Second, by assuming that the first innovator can participate in the second race on the

* Università di Bologna; denicolo@economia.unibo.it.

I am grateful to Luigi Franzoni and seminar participants at Copenhagen, the 1998 ASSET conference on R&D policy, and Università della Calabria for many helpful suggestions. The article has benefited enormously from the incisive comments and detailed instructions of two anonymous referees and Editor Lars Stole. Financial support from MURST is gratefully acknowledged.

same footing as any outsider, I allow for repeated innovation by the same firm.[1] This latter assumption reduces the risk that strong forward protection may impede later improvements; nevertheless, I show that with patent races, weak forward protection can be preferable.

The patent system provides forward protection essentially in two ways (see Scotchmer (1991) and O'Donoghue (1998) for a careful discussion). First, some protection is implicit in the novelty and nonobviousness requirements that any patent application must meet. Second, even a patentable improvement may infringe on the first patent, depending on the latter's leading breadth. These two tools—the novelty requirement and leading breadth—determine the incentives for both generations of research, because they determine the distribution of profit created by second-generation innovations between the firms playing a part in the discovery.

With two tools, there are in principle four patent regimes, according to whether the second innovation is unpatentable and infringing (UI), unpatentable and not infringing (UN), patentable and infringing (PI), and patentable and not infringing (PN). The UN regime can be simply discarded, as no one has any incentive to invest in the second innovation (which, once discovered, would be freely available to all).[2] When the second innovation is unpatentable and infringes, only the first patentholder can lawfully use it, so he alone has an incentive to invest in the improvement. Thus in regime UI the intertemporal externality is fully internalized. In regime PI, the second patent infringes, so the two patents are reciprocally blocking, and neither innovator can use the improvement without the other's consent. This allows the first innovator to collect some rent from the second innovation even if this is made by an outsider. When the second patent does not infringe, the first patentee has no control rights on the second innovation, and therefore in regime PN he collects none of the second-stage rent. Clearly, the three regimes are in decreasing order of forward protection: UI provides the strongest protection to the first innovator, PN the weakest.

The economic effect of moving from UI to PI or from PI to PN is to decrease the returns to the first innovation and increase those to the second (Proposition 1). Such a movement would never be socially desirable if the only source of inefficiency in the market equilibrium were the intertemporal externality due to the sequential nature of the innovations. In general, however, there are two additional sources of inefficiency: with patent races, the winner-takes-all aspect tends to lead to overinvestment because of inefficient competitive externality; on the other hand, because the private returns from successful innovation fall short of the social returns, there is a countervailing tendency to underinvest. In the presence of these additional effects, the policy problem no longer amounts to correcting the intertemporal externality.

I show that regime UI always leads to underinvestment in the second innovation (Proposition 2), as the existence of R&D competition in the first stage of research tends per se to counteract the intertemporal externality, thus making full internalization suboptimal. Although the movement from UI to PI or PN involves a discrete rather than a marginal change in R&D investment and so need not be socially desirable, there is clearly room for a welfare improvement. Indeed, one of the findings of this article is

[1] See also O'Donoghue (1998), who however models a long sequence of innovations. As he observes, with a long sequence of innovations all innovations are alike (each builds on the previous one and opens the way to the subsequent improvement), so the distinction between basic innovations and improvements is lost.

[2] To simplify, I assume here that in the absence of patent protection one cannot make a profit out of an innovation, although I recognize that in practice trade secrets, lead time, and so forth may permit the appropriation of some of the innovation's value even with no patent protection.

that with symmetric innovations, social welfare is highest under regime PN and lowest under UI (Proposition 3).

Relaxing the symmetry assumption, I identify various factors that militate against strong forward protection (Propositions 4 and 5). First, with free entry into R&D activity profits are dissipated, so that expected social welfare is entirely determined by the nonappropriable value of the innovations. Consequently, if the nonappropriable value of the second innovation is large compared to that of the first, it becomes socially valuable to shift incentives from the original innovation to the improvement by reducing forward protection. Second, if the first innovation becomes more profitable, then private benefits will provide stronger motivation for investment in the first innovation; transferring rents to the first-stage innovator becomes less attractive. Hence, as the first innovation's profitability increases, we should expect PN to be more attractive than PI, and PI to be more attractive than UI. A similar argument explains why strong forward protection becomes less desirable as the first innovation becomes more obvious (i.e., the unit R&D cost in the first race decreases). Conversely, forward protection becomes more desirable as the nonappropriable value of the first innovation goes up, the profitability of the second innovation increases, or the second innovation becomes easier to obtain. Finally, if business stealing is introduced, then under PN some of the first-period rents are now transferred to the second-stage innovator. Since PN is already biased in favor of the latter, this makes PI relatively more attractive (Proposition 6).

After presenting the basic model and discussing the patent regimes in Section 2, I proceed to the welfare comparison in Section 3. Section 4 analyzes the case of business stealing, and Section 5 concludes. All proofs are in the Appendix.

2. The model

■ I model two patent races that come in a sequence in the sense that only after the first innovation is achieved can the race for the second begin. Apart from any barrier created by the patent system itself, there is free entry in each race.

I do not rely on any specific model of the innovation and of the downstream product market. These aspects are black-boxed and summarized by the parameters measuring the social and private returns from innovation. I distinguish between two cases, according to whether successive innovations are substitutes or not. I assume initially that the second innovation does not affect the profitability of the first (no business stealing). One can think of the second innovation as an application of the original invention to a different market, like in Matutes, Regibeau, and Rockett (1996). When the second-generation product is a substitute for the first one, by contrast, there may be business stealing. This case is analyzed in Section 4. Let V_1 denote the flow of profit accruing to the first inventor. Then, assuming for simplicity an infinite patent life, the private value of the first innovation is $v_1 = V_1/r$, where r is the discount rate. Likewise, let $v_2 = V_2/r$ denote the private value of the second innovation.

Following Loury (1979) and Dasgupta and Stiglitz (1980), I assume that at the beginning of each race $t = 1, 2$, each participating firm i decides its R&D effort x_{it} and pays a lump-sum cost $c_t x_{it}$, where c_t is the unit R&D cost in the tth race. The R&D effort determines the expected time to successful completion of the project according to a Poisson discovery process. The Poisson assumption implies that there is no cumulative learning (see Reinganum (1989) for a discussion) but is easy to analyze. The research projects of different firms are taken to be independent of one another, so that the aggregate instantaneous probability of success is simply the sum of the individual probabilities. To get a closed-form solution, I assume a linear hazard function,[3] which implies constant returns to scale in R&D.[4]

I focus on the model's subgame-perfect Nash equilibria. Thus, the solution is worked out backward, starting from the second race.

□ **The second patent race.** Under regime PN, the first innovator has exclusive rights to the first innovation but no control rights over the second innovation. Thus there is free entry in the second patent race, and the second innovator does not have to share his profit v_2 with the first. At the beginning of the second race, the payoff function of a generic firm i (i.e., the present value of expected profit, net of R&D costs) is

$$\pi_i = \int_0^\infty e^{-(X_2+r)\tau} x_{i2} v_2 \, d\tau - c_2 x_{i2} = \frac{x_{i2} v_2}{X_2 + r} - c_2 x_{i2}, \tag{1}$$

where $X_2 = \Sigma_i \, x_{i2}$ is the aggregate R&D effort. With free entry into the R&D industry, X_2 is determined by the zero-profit condition $\pi_i = 0$, yielding (with obvious notation)

$$X_2^{PN} = \frac{v_2}{c_2} - r. \tag{2}$$

Next consider regime UI, where the second-generation technology is not patentable. Here only the first innovator would have any incentive to develop the new technology, as he could prevent any other firm from using the improvement and could himself use it without paying any royalty. (I discuss this assumption at greater length in the concluding section.) Thus, the winner of the first race (the leader) is effectively a monopolist at the second stage. The leader chooses his R&D investment to maximize expected profit (1), yielding

$$X_2^{UI} = \sqrt{\frac{r v_2}{c_2}} - r. \tag{3}$$

Thanks to his monopoly, the leader reaps a positive extra-profit in the second race, $\pi_L^{UI} = (\sqrt{v_2} - \sqrt{c_2 r})^2$.

Finally, under regime PI the second innovation is patentable but there is infringement. In this case, the second patentee can prevent all others (including the leader) from using his improvement but cannot himself exploit it without a license from the first patentee. And while the first patentee can obviously continue to use the original device, he cannot use the improvement without the consent of the second. There must therefore be bargaining between the two patentees, and the private returns v_2 from the second innovation will be shared. Since neither firm can use the improvement without

[3] With a linear hazard function, it is convenient to postulate that R&D costs are paid at the beginning of the race because the firm's maximization problem may then have an interior solution. If instead R&D costs are assumed to be flow costs that are paid until the innovation occurs, as in Lee and Wilde (1980), a concave hazard function (at least over the relevant range) would be required to get an interior solution. For a comparison of these two approaches, see Reinganum (1989).

[4] Ignoring the integer constraint, this is consistent with Dasgupta and Stiglitz's assumption that each R&D project is characterized by an optimal scale, but that each firm can run many R&D projects simultaneously. These assumptions eliminate duplication of effort. However, this possibility can be reobtained by adding a fixed R&D cost to be paid by any active research firm—see footnote 11 below.

the consent of the other, it is plausible to posit that the bargaining process will lead to a fifty-fifty split.

To solve for the R&D equilibrium in regime PI, let x_2^L denote the leader's second-period investment, and X_2^O the aggregate investment by outsiders (i.e., all other firms). Because there is free entry, it must be that

$$\frac{v_2}{2} - c_2(x_2^L + X_2^O + r) \leq 0, \tag{4}$$

with $X_2^O > 0$ only if equality holds. The leader's profit in the second race is

$$\pi_L = \frac{x_2^L v_2 + \frac{1}{2} X_2^O v_2}{x_2^L + X_2^O + r} - c_2 x_2^L, \tag{5}$$

and the corresponding first-order condition is

$$\frac{\left(\frac{1}{2} X_2^O + r\right) v_2}{c_2} = (x_2^L + X_2^O + r)^2. \tag{6}$$

Assuming $X_2^O > 0$, one can solve equations (4) and (6) to get $x_2^L = r$ and $X_2^O = v_2/2c_2 - 2r$, so that aggregate R&D investment is

$$X_2^{PI} = x_2^L + X_2^O = \frac{v_2}{2c_2} - r. \tag{7}$$

Substituting this into (5) gives $\pi_L^{PI} = v_2/2 - c_2 r$. Clearly,

$$\pi_L^{UI} - \pi_L^{PI} = \frac{1}{2}(\sqrt{v_2} - 2\sqrt{c_2 r})^2 > 0.$$

The above solution holds for $X_2^O > 0$, i.e., for $v_2/c_2 > 4r$. When $v_2/c_2 \leq 4r$, equilibrium entails $X_2^O = 0$ and the leader behaves in the stand-alone optimal way, so that the equilibrium R&D investment is the same as under regime UI.

This is the solution to a simultaneous-move game between the first-stage patentee and the outsiders. The same aggregate outcome, however, would obtain in a Stackelberg game in which the leader is able to choose second-period investment publicly before any other firm.[5]

□ **The first patent race.** Having found the equilibrium in the second race, I now turn to the first race. There is always free entry in the first race, but the prize to the winner differs across regimes. Under regime PN, the prize simply equals v_1, as the first

[5] This follows because the zero-profit condition determines the outsiders' aggregate best-response function; since the best-response function has slope −1, the aggregate level of investment is unchanged by the game's timing. In the Stackelberg equilibrium, it turns out that the leader's profit depends only on aggregate investment, and it is therefore a matter of indifference whether research is carried out by the leader or by the outsiders.

patentee collects no rent from the second innovation. Thus the free-entry equilibrium condition gives

$$X_1^{PN} = \frac{v_1}{c_1} - r. \tag{8}$$

Under PI, the prize in the first race is $v_1 + \pi_L^{PI}$, yielding (for $v_2/c_2 \geq 4r$)

$$X_1^{PI} = \frac{v_1 + \frac{1}{2}v_2}{c_1} - \left(1 + \frac{c_2}{c_1}\right)r. \tag{9}$$

Finally, under UI (and under PI for $v_2/c_2 \leq 4r$), the prize is $v_1 + \pi_L^{UI}$, and one gets

$$X_1^{UI} = \frac{v_1 + (\sqrt{v_2} - \sqrt{c_2 r})^2}{c_1} - r. \tag{10}$$

To avoid corner solutions, I assume $v_1/c_1 > r$ and $v_2/c_2 > r$, which guarantees that R&D investment is positive in all cases.

□ **Comparative statics.** I can now compare aggregate R&D investment across regimes. To facilitate the comparison, I present a general solution that encompasses the three regimes as special cases. Let the parameter α denote the share of second-stage rent that goes to the first innovator when another firm wins the second race. Proceeding as in the analysis of regime PI, we obtain

$$X_2(\alpha) = \frac{(1 - \alpha)v_2}{c_2} - r \tag{11}$$

and $\pi_L(\alpha) = \alpha v_2 - [\alpha/(1 - \alpha)]c_2 r$, whence we obtain

$$X_1(\alpha) = \frac{v_1 + \alpha v_2}{c_1} - \frac{\alpha}{1 - \alpha}\frac{c_2}{c_1}r - r. \tag{12}$$

Clearly, regime PN corresponds to $\alpha = 0$. When $\alpha \geq \hat{\alpha}$, where $\hat{\alpha}$ is implicitly defined as the solution to $(1 - \alpha)^2 = rc_2/v_2$, a successful outsider would obtain so low a share of second-period rents that entry in the second race is effectively blockaded; thus, $\alpha = \hat{\alpha}$ corresponds to regime UI. Regime PI results from setting α equal either to ½ or to $\hat{\alpha}$, whichever is lower.

The parameter α may reflect the first innovator's bargaining power, but it is difficult to find a natural interpretation of a marginal change in α in terms of patent policy. Nevertheless, it helps compare the three patent regimes to regard α as a continuous index of the strength of forward protection.

As one should expect, when stronger protection is provided to the first inventor, investment in the first innovation is stimulated, that in the second innovation discouraged.

Proposition 1. R&D investment in the first (second) innovation is increasing (decreasing) in the degree of forward protection α.

This means that $X_1^{UI} \geq X_1^{PI} > X_1^{PN}$ and $X_2^{PN} > X_2^{PI} \geq X_2^{UI}$, where all inequalities are strict if $v_2/c_2 > 4r$.

3. Welfare analysis

■ For a variety of reasons the social returns from innovation may be greater than the private returns. Let $v_2 + s_2$ be the social value of the second innovation, and $v_1 + s_1$ the stand-alone social value of the first innovation, with $s_1, s_2 \geq 0$. (As Green and Scotchmer (1995) have pointed out, the social value of the first innovation includes the option value of investing to obtain the second innovation, $\max_{X_2} [X_2/(X_2+r)(v_2 + s_2) - c_2X_2]$.) Expected social welfare, evaluated at the beginning of the first patent race, is therefore

$$W(X_1, X_2) = P(X_1)\{(v_1 + s_1) + [P(X_2)(v_2 + s_2) - c_2X_2]\} - c_1X_1, \tag{13}$$

where $P(X_t) \equiv X_t/(X_t + r)$ may be interpreted as the ''adjusted probability'' of innovating (with a Poisson discovery process, the innovation eventually occurs with probability one, but since there is discounting, a delayed success is valued less than instant success).[6]

From (13), one can easily calculate the socially optimal levels of R&D investment, X_1^S and X_2^S:

$$X_1^S = \sqrt{\frac{r}{c_1}[(v_1 + s_1) + (\sqrt{(v_2 + s_2)} - \sqrt{c_2 r})^2]} - r \tag{14}$$

$$X_2^S = \sqrt{\frac{r(v_2 + s_2)}{c_2}} - r. \tag{15}$$

As is well known, there are two reasons why equilibrium investment may differ from the optimum. First, the private returns from innovation may be lower than the social returns, and this leads to underinvestment in R&D. The higher s_1 and s_2, the stronger this effect. Here, underinvestment in the first race may also be caused by the intertemporal externality that arises when the first innovator fails to appropriate the option value of investing in the second innovation. Second, because in a patent race only the successful contestant gains, there is a tendency to overinvest as firms strive to be first. This winner-takes-all effect is at work in the first race under all regimes, but under regime UI it vanishes in the second race. The intertemporal externality is fully (partially) internalized under regime UI (PI); there is no internalization in PN.

In the market equilibrium, expected profits are bid to zero at the beginning of the first race, and in regimes PI and PN, outsiders' expected profits are bid to zero at the beginning of the second race. Therefore, under all patent regimes social welfare reduces to

$$W = P(X_1)[s_1 + P(X_2)s_2]. \tag{16}$$

[6] With this interpretation of $P(X)$, it should be clear that the model can be extended to the case where innovations are instantaneous and R&D effort determines the probability rather than the timing of success. While some of my results depend on the specific functional form of $P(X)$, others are driven by the simple fact that $P(X)$ is bounded above by one, implying that an increase in X must eventually have a negligible impact on P.

Equation (16) tells us that expected social welfare is equal to the nonappropriable value of each innovation, s_1 and s_2, multiplied by their respective ‘‘adjusted probabilities’’ of discovery, i.e., $P(X_1)$ for the first innovation and $P(X_1)P(X_2)$ for the second. We know from Proposition 1 that $P(X_1)$ is increasing and $P(X_2)$ is decreasing in the degree of forward protection α. Strengthening forward protection is beneficial only if the first-stage benefits outweigh the second-stage losses.

Lemma 1. Social welfare is strictly concave in the degree of forward protection: $W''(\alpha) < 0$.

This lemma follows from the concavity of $P(X)$. Since Lemma 1 means that $W(\alpha)$ is single peaked, it cannot happen that both $\alpha = 0$ and $\alpha = \hat{\alpha}$ are preferable to intermediate degrees of forward protection. Consequently, it is never optimal to jump from PN to UI as any one of the model's parameters changes continuously (unless, of course, PI and UI coincide).

Proposition 2. If the nonappropriable value of the second innovation is positive ($s_2 > 0$), in regime UI a marginal increase in R&D investment in the second innovation, X_2, is always welfare improving.

The reason is that in this regime there is always underinvestment in the second race with respect to the social optimum (i.e., $X_2^{UI} < X_2^S$) because of the absence of R&D competition. Thus, a marginal increase in X_2 has a first-order positive effect on social welfare. The increase in X_2 diminishes the leader's profit, and hence first-stage investment; but as the former is maximized at X_2^{UI}, this effect (which may further increase or decrease social welfare depending on whether there is over- or underinvestment in the first race) is second order. But since a change in regime involves a discrete rather than a marginal change in R&D investment, Proposition 2 does not mean that regime UI is dominated by regime PI.

I next focus on the benchmark case of symmetric innovations, i.e., $v_1 = v_2$, $c_1 = c_2$ and $s_1 = s_2$.

Proposition 3. In the symmetric case, social welfare is highest in regime PN and lowest in regime UI, i.e., $W^{PN} > W^{PI} \geq W^{UI}$, where the last inequality is strict if $v_2/c_2 > 4r$.

Although this result rests on the specific functional form $P(X) = X/(X + r)$, it nicely illustrates one theme of this article, namely that broad forward protection may be costly in the presence of R&D competition.

Relaxing the symmetry assumption, we can identify three key determinants of the welfare ranking of the regimes: the ratio between the nonappropriable values of the two innovations, $\sigma = s_2/s_1$ (since social welfare is homogeneous of degree one in s_1 and s_2, only σ matters); the unit R&D costs, c_1 and c_2, which may be thought of as indices of the nonobviousness of the innovations; and the profitability of the innovations, v_1 and v_2. In view of Lemma 1, I compare regimes PN and PI, then PI and UI. (However, Proposition 4 would continue to hold with W^{UI} replacing W^{PI}.)

Proposition 4. The ratio W^{PN}/W^{PI} is (i) increasing in the ratio $\sigma = s_2/s_1$, (ii) increasing in v_1, (iii) decreasing in v_2, (iv) decreasing in c_1, and (v) increasing in c_2.

Some comments are in order. First of all, it is clear that as σ increases it is socially beneficial to shift incentives from the first to the second innovation, given that social welfare is entirely determined by the nonappropriable value of the two innovations. Some intuitive insight into the subsequent results may be gained by noting that $P(X)$ is bounded above by one, so that regardless of the protection regime, when investment in either one of the two innovations is so great that the ‘‘adjusted probability’’ of success is close to one, the welfare ranking is determined primarily by the timing of the other innovation. Thus, when the first (second) innovation becomes more profitable, investment in the first (second) innovation is more strongly encouraged by the prospect of commercial success, and transferring rents to the first patentee becomes less (more) socially beneficial. Analogous arguments explain the effect of a change in the nonobviousness of the two innovations as measured by their respective R&D costs, c_1 and c_2.

Let us now compare regimes PI and UI. Since they lead to the same R&D investment if $v_2/c_2 < 4r$, I focus on the case $v_2/c_2 \geq 4r$.

Proposition 5. The ratio W^{PI}/W^{UI} is (i) increasing in the ratio $\sigma = s_2/s_1$, (ii) increasing in v_1, and (iii) decreasing in c_1.

Results (i)–(iii) are the same as in Proposition 4. The ratio W^{PI}/W^{UI}, however, is not monotonic in v_2 and c_2. To see this, note that $W^{PI}/W^{UI} = 1$ for $c_2 = v_2/4r$. If c_2, say, is now slightly reduced, X_2^{PI} will become slightly greater than X_2^{UI}, which by Proposition 2 implies that $W^{PI} > W^{UI}$. As c_2 is further reduced, however, $P(X_2)$ tends to one in both regimes and therefore the welfare ranking is driven by the fact that $X_1^{UI} > X_1^{PI}$, implying that $W^{PI} < W^{UI}$. This means that W^{PI}/W^{UI} first rises and then falls as c_2 decreases. The intuition for this nonmonotonicity is that when c_2 falls (or v_2 rises) not only are the incentives to invest in the second innovation higher, but the rents transferred to the first patentee also increase. When X_2^{PI} is close to X_2^{UI}, by Proposition 2 the latter effect must prevail, but the former eventually dominates. Despite this nonmonotonicity in the ratio W^{PI}/W^{UI}, the previous arguments show that the improvement should be patentable when its R&D cost is high (or its private value is low) and should be unpatentable when c_2 is low.

To conclude this section I summarize the main results and briefly point out their policy implications. First, weak forward protection is desirable under a broad variety of circumstances. Second, it is important to distinguish between the appropriable and nonappropriable value of the innovations, as they have opposite policy implications: due to the winner-takes-all effect, with patent races there is lesser need to protect highly profitable innovations, whereas strong protection is desirable when a large fraction of the returns to innovation are nonappropriable. Consequently, weak forward protection is desirable when the nonappropriable value of the second innovation is relatively high, while infringement should be found when the private returns to the second (first) innovation are high (low).[7] Finally, a relatively nonobvious improvement should be patentable and not infringe.

4. Business stealing

■ So far I have assumed that the second-generation innovation does not affect the profitability of the first but simply adds a new flow of profit V_2. If the first- and second-generation products are substitutes, however, there will normally be business stealing. In this section I assume that the first inventor gets a profit flow equal to V_1 until the second innovation occurs; thereafter, the second innovator gets a profit of $\mu V_1 + V_2$, and the profit to the first inventor falls to $(1 - \mu)V_1$. The total (private and social) value of the two innovations is as in the previous sections; however, the distribution

[7] This result runs counter to the suggestion of Merges and Nelson (1990) that courts should find infringement only when the value of the improvement is low compared to the value of the original invention. It also differs from Chang's (1995) finding that the optimal infringement rule is nonmonotonic.

of profit is different.[8] When $\mu = 0$, we are back to the model of the previous sections; when $\mu = 1$, we have the extreme case in which the second innovation drives the profit from the first to zero.

Obviously, the division of profit is irrelevant if the second innovation is not patentable and the first inventor obtains a research monopoly, or if there is infringement and the second innovation is exploited cooperatively by the two patentees. (The disagreement point in the bargaining process is independent of whether there is business stealing or not.) However, business stealing does change the equilibrium under regime PN. In this case, equilibrium R&D effort in the second race is

$$X_2^{PN} = \frac{\mu v_1 + v_2}{c_2} - r. \tag{17}$$

The prize to the first innovator is no longer given by v_1; instead, it is $(1 - \mu)v_1$ plus the profit flow μV_1 discounted at rate r until the occurrence of the second innovation, that is,

$$(1 - \mu)v_1 + \frac{\mu V_1}{X_2^{PN} + r} = (1 - \mu)v_1 + rc_2\frac{\mu v_1}{\mu v_1 + v_2}. \tag{18}$$

It follows that in the first race, equilibrium R&D investment will be[9]

$$X_1^{NP} = (1 - \mu)\frac{v_1}{c_1} + \left(\frac{c_2}{c_1}\frac{\mu v_1}{\mu v_1 + v_2} - 1\right)r. \tag{19}$$

Clearly, the presence of business stealing reduces investment in the first race and increases it in the second. As such, business stealing is equivalent to negative forward protection ($\alpha = -\mu v_1/v_2$). Although the ratio W^{PN}/W^{PI} need not be monotonic in μ, the concavity of $W(\alpha)$ implies that the welfare ranking of PI and PN is monotonic in a sense that is made precise in the following proposition.

Proposition 6. If $W^{PI} \geq W^{PN}$ with no business stealing, then $W^{PI} > W^{PN}$ for any positive degree of business stealing.

Proposition 6 tells us that more business stealing makes the infringement regime more attractive. The economic intuition underlying this proposition is that the presence of business stealing means that regime PN is even more heavily biased in favor of the second-generation innovation. Since there are decreasing returns to stimulating the second innovation because of the concavity of $P(X)$, if it does not pay to move from PI to PN with no business stealing, such a move cannot be welfare improving with business stealing.

[8] Note that I allow for business stealing but not for the erosion of profit that may be associated with product market competition, in contrast with O'Donoghue, Scotchmer, and Thisse (1998). The erosion of profit can be avoided if the patentees collude, but the second patentee can engage in business stealing by threatening to drain away the first innovator's profit by competition.

[9] Note that when μ is close to one, this expression may become negative, in which case there will be no investment to obtain the first innovation and hence no technical progress at all. The reason is that because of the weak forward protection and the presence of business stealing, there is too much investment in the second race. This exacerbates the problem of inducing firms to invest in the first invention, since the second innovation will follow soon and there will thus be little time to recoup R&D expenditures.

5. Conclusion

■ I have reexamined the issue of optimal patent design with cumulative technical progress using a standard model of patent races with a Poisson discovery process. This model has well-known limits but allows one to analyze how the dynamic externality arising from the cumulative nature of innovation interacts with other sources of inefficiency discussed in the patent race literature, namely the winner-takes-all effect and the difference between the private and social returns.

These latter sources of inefficiency have been relatively neglected in the two-stage patent design literature so far, leading to an underestimate of the social costs of strong forward protection. For instance, in all of Green and Scotchmer (1995), Chang (1995), and Scotchmer (1996), if the first innovator could obtain the improvement, regime UI would provide a complete solution to the problem of the division of profit. In Chang (1995), this conclusion is driven by the assumption that the social and private returns from innovation coincide, i.e., $s_1 = s_2 = 0$. The critical assumption in Green and Scotchmer (1995) and Scotchmer (1996) is that *ex ante* agreements are feasible. Under either assumption, there is no underinvestment in the second innovation.

With *ex ante* agreements, efficiency in the development of the improvement is guaranteed even if the first innovator cannot invest in the second innovation, in which case the problem becomes one of transferring as much as possible of the total profit from the two inventions to the original innovator. Although Green and Scotchmer (1995) focus on the comparison between PI and PN, clearly in their model UI would best achieve this aim.[10] Scotchmer (1996) shows that the same is true when two firms are able to obtain the improvement, and the first innovator can contract with both of them. As she aptly observes, however, ‘‘denying second-generation patents could stifle innovation if there were impediments to *ex ante* contracting’’ (p. 330). By focusing on this latter case, I have shown that regime UI generally entails underinvestment in the second innovation, implying that full internalization of the intertemporal externality need not be optimal.

Moreover, in my model the mere existence of R&D competition in the first race tends to counteract the intertemporal externality, and this further weakens the case for strong forward protection. This finding may be contrasted with Kitch’s (1977) early argument that strong forward protection is attractive precisely because it prevents duplication of R&D effort at the development stage. (To capture the notion of inefficiencies from duplication of effort, one should extend the model by adding a fixed cost to be paid by any firm that is active in R&D, as in Denicolò (1999).)[11] In my model, any gain in efficiency that may be obtained by eliminating overinvestment or duplication of effort in the second race is dissipated by R&D competition in the first race.

My model could be extended in several ways, some of which have already been mentioned. One assumption that could be relaxed is that outsiders have no incentive to invest if the second innovation is unpatentable. This presumes that lacking property rights, one cannot make a profit out of an invention. Scotchmer (1996) takes a more optimistic view of the outsider’s ability to appropriate a sizable share of the value of

[10] Because of the subtle interaction between *ex ante* and *ex post* bargaining, in Green and Scotchmer (1995) the welfare ranking of the three regimes need not be single peaked, and so they find that PN might be preferable to PI under certain circumstances even if UI would always be the best of all regimes.

[11] The equilibrium analyzed in the present article is reobtained as the fixed cost goes to zero. It can be shown that increasing the fixed cost is not clearly biased in favor of or against any patent regime. With a positive fixed cost, the free-entry equilibrium involves a finite number of firms, in contrast to the case of constant returns to scale where there is an infinite number of firms each conducting a negligible amount of research.

the second innovation (see also Anton and Yao (1994)). Postulating that a successful outsider can bargain with the first-period patentee without disclosing his invention, then even if only the first-period patentee can use the improvement and the outsider's R&D costs are sunk, a successful outsider has some bargaining power. The disagreement point in the bargaining process is zero for the outsider (he cannot use the innovation without an agreement) and $x_2^L v_2/(x_2^L + r)$ for the leader (i.e., the discounted value of waiting for his own discovery, given that R&D costs are sunk). Since the value of the innovation is v_2, there is a positive surplus to be split, which is given by $rv_2/(x_2^L + r)$. Proceeding as in Section 2, one obtains again $X_1^{UI} \geq X_1^{PI}$ and $X_2^{UI} \leq X_2^{PI}$. However, Proposition 2 no longer holds, in that regime UI no longer protects the leader from outsiders' competition in the second race.

More explicit modelling of the product market may also offer new insights. In particular, when first- and second-generation products may compete in the market, under regime PN there will typically be some profit erosion. My assumption that the total return from the two innovations is fixed may then be inappropriate, unless the patentees can collude. This raises a key antitrust issue, namely whether collusion between patentees who would otherwise compete should be permitted or not. The analysis of this issue in a patent race framework is left for future work.

Appendix

■ The proofs of Propositions 1–6 and Lemma 1 follow.

Proof of Proposition 1. Differentiating (11) and (12) we get

$$\frac{dX_2}{d\alpha} = -\frac{v_2}{c_2} < 0 \tag{A1}$$

and

$$\frac{dX_1}{d\alpha} = \frac{v_2}{c_1} - \frac{1}{(1-\alpha)^2}\frac{c_2}{c_1}r. \tag{A2}$$

The latter derivative is positive for $\alpha < \hat{\alpha}$ *Q.E.D.*

Proof of Lemma 1. Using (11), (12), and (16) we can express social welfare as a function of α, $W(\alpha)$. Twice differentiating this function and using (A1) and (A2) we get

$$\frac{dW}{d\alpha} = P'(X_1)\left[\frac{v_2}{c_1} - \frac{1}{(1-\alpha)^2}\frac{c_2}{c_1}r\right][s_1 + s_2P(X_2)] - s_2\frac{v_2}{c_2}P(X_1)P'(X_2) \tag{A3}$$

and

$$\frac{d^2W}{d\alpha^2} = P''(X_1)\left[\frac{v_2}{c_1} - \frac{1}{(1-\alpha)^2}\frac{c_2}{c_1}r\right]^2[s_1 + s_2P(X_2)] - \frac{2}{(1-\alpha)^3}\frac{c_2}{c_1}rP'(X_1)[s_1 + s_2P(X_2)]$$
$$- 2s_2\frac{v_2}{c_2}P'(X_1)P'(X_2)\left[\frac{v_2}{c_1} - \frac{1}{(1-\alpha)^2}\frac{c_2}{c_1}r\right] + s_2\left(\frac{v_2}{c_2}\right)^2 P(X_1)P''(X_2) < 0,$$

where the inequality follows from $P'(X) > 0$ and $P''(X) < 0$. *Q.E.D.*

Proof of Proposition 2. It suffices to show that $W'(\hat{\alpha}) < 0$. To prove this, note that at $\alpha = \hat{\alpha}$, (A3) reduces to $dW/d\alpha = -s_2v_2P(X_1)P'(X_2)/c_2 < 0$. *Q.E.D.*

Proof of Proposition 3. In view of Lemma 1, it suffices to show that $W^{PN} > W^{PI}$. Using (16), expected social welfare can be calculated as follows:

$$W^{PN} = s_1 \frac{(\eta - 1)(2\eta - 1)}{\eta^2}, \tag{A4}$$

$$W^{PI} = \begin{cases} 2s_1(3\eta - 4)(\eta - 1)/[\eta(3\eta - 2)] & \text{for } v_2/c_2 \geq 4r \quad \text{(A5)} \\ 2s_1(\sqrt{\eta} - 1)(2\sqrt{\eta} - 1)/[\eta + (\sqrt{\eta} - 1)^2](= W^{UI}) & \text{for } v_2/c_2 \leq 4r, \quad \text{(A6)} \end{cases}$$

where $\eta \equiv v_1/c_1 r$ $(= v_2/c_2 r)$. Recall that to avoid corner solutions it is assumed that $\eta > 1$.

When $\eta \leq 4$, from (A4) and (A6) we get

$$\frac{W^{PN}}{W^{PI}} = \frac{4\eta^2\sqrt{\eta} - 4\eta\sqrt{\eta} + 2\eta + \sqrt{\eta} + 1}{4\eta^2\sqrt{\eta} - 2\eta^2}.$$

From this expression, it follows that $W^{PN}/W^{PI} \lesseqqgtr 1$ if and only if $2\eta^2 - 4\eta\sqrt{\eta} + 2\eta + \sqrt{\eta} + 1 \gtreqqless 0$. But we have $2\eta^2 - 4\eta\sqrt{\eta} + 2\eta + \sqrt{\eta} + 1 = (\sqrt{\eta} - 1)[2\eta(\sqrt{\eta} - 1) + 1] > 0$, where the inequality follows from $\eta > 1$. When $\eta \geq 4$, from (A4) and (A5) we get $W^{PN}/W^{PI} = (6\eta^2 - 7\eta + 2)/(6\eta^2 - 8\eta) > 1$, whence the result follows immediately. *Q.E.D.*

Proof of Proposition 4. (i) From (16), we have

$$\frac{W^{PN}}{W^{PI}} = \frac{P(X_1^{PN})}{P(X_1^{PI})} \frac{1 + \sigma P(X_2^{PN})}{1 + \sigma P(X_2^{PI})}. \tag{A7}$$

Differentiating (A7) one gets

$$\frac{d(W^{PN}/W^{PI})}{d\sigma} = \frac{P(X_1^{PN})}{P(X_1^{PI})} \frac{P(X_2^{PN}) - P(X_2^{PI})}{[1 + \sigma P(X_2^{PI})]^2},$$

which by Proposition 1 is positive.

(ii) Using (2), (7), (8), and (9) we can easily check that $dP(X_1)/dv_1 = r/[c_1(X_1 + r)^2]$ and $dP(X_2)/dv_1 = 0$ in both regimes. Then

$$\frac{d(W^{PN}/W^{PI})}{dv_1} = \frac{1 + \sigma P(X_2^{PN})}{1 + \sigma P(X_2^{PI})} \frac{r}{c_1[P(X_1^{PI})]^2} \left[\frac{P(X_1^{PI})}{(X_1^{PN} + r)^2} - \frac{P(X_1^{PN})}{(X_1^{PI} + r)^2} \right].$$

By Proposition 1, the term inside brackets is positive, whence the result follows.

(iii) The proof is similar to the proof of point (ii), and it exploits the fact that $dP(X_1)/dc_1 = -r/[c_1(X_1 + r)]$ and $dP(X_2)/dc_1 = 0$ in both regimes.

(iv) Again from (2), (7), (8), and (9) we obtain

$$dP(X_1^{PN})/dv_2 = 0, \quad dP(X_1^{PI})/dv_2 = r/[2c_1(X_1^{PI} + r)^2], \quad dP(X_2^{PN})/dv_2 = rc_2/v_2^2, \quad \text{and}$$
$$dP(X_2^{PI})/dv_2 = 2rc_2/v_2^2.$$

Differentiating (A7) we then have

$$\frac{d(W^{PN}/W^{PI})}{dv_2} = \frac{P(X_1^{PN})}{P(X_1^{PI})} \frac{\sigma rc_2}{v_2^2[1 + \sigma P(X_2^{PI})]^2} \{1 + \sigma P(X_2^{PI}) - 2[1 + \sigma P(X_2^{PN})]\}$$
$$- \frac{1 + \sigma P(X_2^{PN})}{1 + \sigma P(X_2^{PI})} \frac{P(X_1^{PN})}{[P(X_1^{PI})]^2} \frac{r}{2c_1(X_1^{PI} + r)^2}.$$

Since the term inside curly brackets is negative by Proposition 1, the derivative is negative.

(v) The proof is similar to the proof of point (iv), exploiting the fact that

$$dP(X_1^{PN})/dc_2 = 0, \quad dP(X_1^{PI})/dc_2 = -r^2/[c_1(X_1^{PI} + r)^2], \quad dP(X_2^{PN})/dc_2 = -r/v_2, \quad \text{and}$$
$$dP(X_2^{PI})/dv_2 = -2r/v_2.$$

Q.E.D.

Proof of Proposition 5. The proof is similar to the proof of points (i)–(iii) of Proposition 4 and is not repeated here. *Q.E.D.*

Proof of Proposition 6. Comparison of (17) and (19) with (11) and (12) confirms that they coincide for $\alpha = -\mu v_1/v_2$. The result then follows immediately from Lemma 1. *Q.E.D.*

References

ANTON, J.J. AND YAO, D.A. ''Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights.'' *American Economic Review,* Vol. 84 (1994), pp. 190–209.

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.

DENICOLÒ, V. ''The Optimal Life of a Patent When the Timing of Innovation Is Stochastic.'' *International Journal of Industrial Organization,* Vol. 17 (1999), pp. 827–846.

GREEN, J.R. AND SCOTCHMER, S. ''On the Division of Profit in Sequential Innovation.'' *RAND Journal of Economics,* Vol. 26 (1995), pp. 20–33.

HELLER, M.A. AND EISENBERG, R.S. ''Can Patents Deter Innovation? The Anti-communist in Biomedical Research.'' *Science,* Vol. 280 (1998), pp. 698–701.

KITCH, E.W. ''The Nature and Function of the Patent System.'' *Journal of Law and Economics,* Vol. 20 (1977), pp. 265–290.

LEE, T. AND WILDE, L.L. ''Market Structure and Innovation: A Reformulation.'' *Quarterly Journal of Economics,* Vol. 94 (1980), pp. 429–436.

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.'' *RAND Journal of Economics,* Vol. 27 (1996), pp. 60–83.

MERGES, R.P. AND NELSON, 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.'' *Journal of Economic Behavior and Organization,* Vol. 25 (1994), pp. 1–24.

O'DONOGHUE, T. ''A Patentability Requirement for Sequential Innovations.'' *RAND Journal of Economics,* Vol. 29 (1998), pp. 654–679.

———, SCOTCHMER, S., AND THISSE, J.-F. ''Patent Breadth, Patent Length, and the Pace of Technological Progress.'' *Journal of Economics and Management Strategy,* Vol. 7 (1998), pp. 1–32.

REINGANUM, J.F. ''The Timing of Innovation: Research, Development, and Diffusion.'' In R. Schmalensee and R.D. Willig, eds., *The Handbook of Industrial Organization.* New York: North-Holland, 1989.

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

———. ''Protecting Early Innovators: Should Second-Generation Products Be Patentable?'' *RAND Journal of Economics,* Vol. 27 (1996), pp. 322–331.