International Journal of Industrial Organization
20 (2002) 801–827

International Journal of Industrial Organization

www.elsevier.com/locate/econbase

How should forward patent protection be provided?

Vincenzo Denicolò[a,*], Piercarlo Zanchettin[b]

[a]*Dipartimento di Scienze Economiche, Università di Bologna, Piazza Scaravilli 2, I-40126 Bologna, Italy*

[b]*Dipartimento di Economia Politica, Università di Siena, Piazza San Francesco 7, Siena, Italy*

Received 5 May 1999; received in revised form 27 October 2000; accepted 23 January 2001

Abstract

We study the optimal combination of the novelty requirement and leading breadth to provide forward patent protection in a simple model of sequential innovation. Forward protection may be provided by a blocking effect, a profit-sharing effect, or both. The novelty requirement and leading breadth are perfect substitutes for low levels of forward protection in that both protect the original innovator by blocking subsequent improvements. As forward protection strengthens, the only appropriate policy variable is leading breadth because it is the most efficient and in some cases the only way to guarantee profit sharing. Under certain conditions, a positive novelty requirement (with complete leading breadth) becomes necessary when the optimal level of forward protection goes still higher.

JEL classification: L00; O31; O34

Keywords: Cumulative innovation; Patent; Novelty requirement; Patent breadth

1. Introduction

When innovation is cumulative, basic innovations need *forward* patent protection, because otherwise future innovators could compete away the original innovators' profits; and because the first innovators should be rewarded for

*Corresponding author. Tel.: +39-051-209-8026; fax: +39-051-221-968.
E-mail address: denicolo@economia.unibo.it (V. Denicolò).

0167-7187/02/$ – see front matter
PII: S0167-7187(01)00080-7

opening the way to the subsequent improvements (Scotchmer, 1991). The patent system provides forward protection in two ways: first, any patent application must meet certain *novelty requirements*; second, even patentable improvements may constitute infringement on the original patent, depending on the first-generation patent's *leading breadth*. To date, the literature on optimal patent design for cumulative innovation has focused almost exclusively on the optimal level of forward protection, dealing with each instrument separately.[1] In a seminal article, Green and Scotchmer (1995) assume that the second innovation is always patentable, and ask whether or not the courts should find an infringement. The issue has also been addressed by Chang (1995), Matutes et al. (1996), O'Donoghue et al. (1998), and Bessen and Maskin (2000) in different frameworks. Other papers, notably van Dijk (1995), Scotchmer (1996), O'Donoghue (1998), and Hunt (1999), focus on the novelty requirement.[2] In this paper, we ask under what conditions the novelty requirement and leading breadth are perfect substitutes, and, when they are not, we further ask how forward protection should be provided: through the novelty requirement, leading breadth, or a combination of the two?

Indeed, the novelty requirement and leading breadth protect early innovators in different ways. The fact that an innovation must satisfy novelty requirements may block or impede second-generation improvements, thus lengthening the first-generation innovators' monopoly. On the other hand, by finding that an improvement infringes on a basic patent, the courts can force the patentees to bargain over profit shares, thus allowing the original innovator to capture some of the rents from the improvement. This means that the novelty requirement has a ''blocking effect,'' and leading breadth a ''sharing effect.'' We argue that sharing is generally better than blocking, because it does not stifle technical progress, and because the intertemporal externality from the cumulative nature of innovation is at least partially internalized. As a consequence, leading breadth tends to dominate the novelty requirement as a way of providing forward protection. Yet, this too needs to be qualified.

First, leading breadth also has a blocking effect: if the second innovation is small and infringes, the profit left to the second innovator may be too low for investment to be profitable (R&D costs are sunk when bargaining between the patent holders takes place). This means that leading breadth inevitably prevents

[1] See Scotchmer (1999) for a survey.

[2] O'Donoghue (1998) briefly analyzes the interaction between the two instruments. He argues that when the size of innovations is endogenous, the novelty requirement forces firms to target larger innovations, thus delaying the next innovation, and shows that such a blocking mechanism can increase forward patent protection even starting from complete leading breadth (see also O'Donoghue and Zweimuller, 1998). However, he does not characterize the optimal policy mix. Denicolò (2000) also considers both instruments simultaneously, but in his model the two patent policy tools have qualitatively similar effects and differ only in the level of forward protection they provide.

some second-generation improvements. In fact, we show that while this effect is always at work, the sharing effect operates only if leading breadth is sufficiently great. Consequently, when forward protection is weak, the novelty requirement and leading breadth are perfect substitutes. By contrast, when forward protection is strong, it is preferable to use leading breadth only. In all cases, however, in our basic model leading breadth is all that is needed to achieve the (second-best) optimum (Proposition 1). We also determine (Proposition 2) when weak or strong forward protection is desirable.

Second, the novelty requirement has no blocking effect when the original innovator can obtain the second-generation improvement. In this case, the optimal patent policy involves no effective leading breadth.

Finally, the novelty requirement may also produce a sharing effect because of the legal doctrine of ''prosecution history estoppel,'' under which all the techniques that have been invented by the second innovator but fail to meet the novelty requirement must be devoted to the public. When there is positive leading breadth, this strengthens the first innovator's bargaining position and increases his share of second-generation profits. Under prosecution history estoppel, we show that when optimal forward protection is sufficiently strong, both instruments must be used (Proposition 3). A similar result holds if firms invest in R&D before the uncertainty over the value of innovations is resolved.

The fact that leading breadth has a blocking effect in addition to its sharing effect carries another interesting implication. In the basic model, we assume that licensing agreements are permitted only in cases of infringement, as under current antitrust policy. If licensing agreements were unrestricted, however, even trifling improvements could become privately valuable, as they could be used as a bargaining weapon to appropriate some of the first innovator's profit. Blocking such trifling improvements is now desirable but can still be obtained using only leading breadth.

The rest of the paper is organized as follows. Section 2 presents the basic model. In Section 3, we characterize optimal patent policy in the basic model, showing that the optimal mix does not involve the novelty requirement. In Section 4 we analyze the case in which collusion is permitted independent of patent rights: again, the optimal policy does not require the novelty requirement. Section 5 presents some extensions entailing a more important role for novelty. Section 6 concludes the paper.

2. The assumptions

2.1. Technology, tastes, timing

Our model is an adapted version of Green and Scotchmer (1995) and Chang

(1995).[3] There are two firms, 1 and 2, and two periods, with zero discount rate. In each period, one firm can invest to obtain an innovation: firm 1 alone can innovate in the first period, firm 2 in the second (the case where firm 1 can innovate repeatedly will be considered later). The first innovation creates a new product of quality q_1, and firm 2 can improve on it by raising the quality level to $q_1 + q_2$ (the case of cost-reducing innovations is similar). Innovation is cumulative: i.e., firm 2 cannot innovate until after firm 1 has made the initial innovation.

Once an innovation is obtained, the good can be produced at zero marginal cost. Consumers, whose mass is normalized to 1, are identical and buy at most one unit of the good. A consumer's utility from consuming one unit of the good of quality q is $q - p$, where p is the price. With this inelastic demand, monopoly pricing entails no deadweight loss.[4]

The size of innovations is exogenous and stochastic. When the firm makes its investment decision, the uncertainty over the size of its innovation has already been resolved (the case in which firms invest before uncertainty is resolved is considered in Section 5). Firm 1, however, faces uncertainty over the value of the second innovation, while the policy-maker faces uncertainty over the value of both innovations. The size of the first innovation, q_1, is drawn from a probability distribution with support $[0, \bar{q}_1]$ and cumulative distribution function $F_1(q_1)$ with $F_1'(q_1) \equiv f_1(q_1)$; likewise, q_2 is drawn from a probability distribution $F_2(q_2)$ with support $[0, \bar{q}_2]$ and $F_2'(q_2) \equiv f_2(q_2)$. Both distribution functions are common knowledge.

The timing of the R&D game is as follows. At the beginning of the first period, firm 1 observes q_1 and decides whether or not to invest in R&D. If it decides not to invest, the game ends. If it does invest, it obtains the first innovation with probability 1 at a deterministic cost c_1. At the beginning of the second period, firm 2 observes the size of the second innovation and decides whether or not to invest. If it invests, it obtains the second innovation with probability 1 at a deterministic cost c_2. To avoid trivial cases where it is never profitable to invest in R&D, we assume that $0 \leq c_1 < 2\bar{q}_1$ and $0 \leq c_2 < \bar{q}_2$.

2.2. Patent policy

We posit that the second innovation is patentable if and only if the improvement, q_2, is at least as large as a critical value, H, that represents the novelty requirement, and that the improvement infringes on the patent if and only if q_2 is

[3] Besides considering both the novelty requirement and leading breadth, the main difference from Green and Scotchmer (1995) is that they allow for ex ante agreements (see footnote 14). Our model is closest to Chang (1995), differing only in our assumptions that R&D costs are deterministic.

[4] The role of this simplifying assumption will be discussed later.

lower than another critical value, K, representing the first patent's leading breadth.[5] The policymaker chooses the novelty requirement, H, and leading breadth, K, before uncertainty is resolved and firms invest in R&D. We assume that the first innovation is always patentable.

Since firm 2 would hardly invest to obtain an unpatentable innovation, whether or not it is an infringement,[6] there is no loss of generality in assuming that $K \geqslant H$; we shall sometimes refer to $\Delta \equiv K - H$ as the ''effective'' leading breadth. A second-generation innovation that is greater than H but less than K is patentable but infringes on the first patent. In this case, while the holder of the first patent can obviously continue to use the first invention, neither patentee can use the improvement without the other's consent. Some form of licensing is inevitable if the second innovation is to be exploited, and the courts and antitrust authorities generally agree that combinations of such ''blocking'' patents are lawful.[7]

Concerning patent length, we suppose that patents last $1 + \tau$ periods, where $0 < \tau \leqslant 1$. Inequality $\tau > 0$ is required to keep the analysis meaningful: if the first patent expires before the second innovation, the novelty requirement and leading breadth can play no role.[8]

2.3. Antitrust policy

When, given patent rights, licensing is not required, the courts and antitrust authorities may prohibit collusion between the two innovators. If collusion is prohibited, as is the case under current antitrust policy, the arrival of the second innovation entails competition and hence profit erosion. Initially we examine this case. Later we shall analyze an alternative and more lenient antitrust policy in which collusive agreements are permitted even if the second innovation does not infringe.

[5] Our modeling strategy closely follows O'Donoghue et al. (1998) and O'Donoghue (1998), who provide a more detailed discussion of relevant aspects of patent law. Note that we are implicitly assuming that the optimal infringement and patentability rules are monotonic in the value of the second innovation; Chang (1995) shows that the optimal infringement rule is, indeed, monotonic in q_2 in his model.

[6] We assume that in the absence of patent protection one cannot make a profit out of an innovation, although of course in practice trade secrets, lead time, etc. may permit appropriation of some of the innovation's value even without protection.

[7] We rule out ex ante agreements, i.e. agreements reached before firm 2's investment is sunk; the optimal policy with ex ante agreements is characterized in footnote 14.

[8] While it is clear that the second innovation cannot infringe on an expired patent, the novelty requirement must be met whether or not the first patent has expired, because the first innovation has become part of the ''prior art.'' However, since the first innovator does not gain from forward protection if his patent has expired, the optimal novelty requirement trivially vanishes if $\tau \leqslant 0$.

2.4. Product market equilibrium

In the first period, given the inelastic demand, firm 1 will price at $p_1 = q_1$ and will gain profit q_1. If firm 2 does not innovate, the same outcome obtains in the second period until the patent expires, when the innovation becomes public and the price falls to zero.

If firm 2 does innovate, we must distinguish between two cases: $q_2 < K$ and $q_2 \geqslant K$. If $H \leqslant q_2 < K$, the two patentees can lawfully enter licensing agreements. (If $q_2 < H$, firm 2 will not innovate.) Before the patent on the first innovation expires, bargaining results in a price equal to $q_1 + \gamma q_2$, with $0 \leqslant \gamma \leqslant 1$. The parameter γ may be seen as an index of the amount of transaction costs in bargaining. If bargaining is efficient ($\gamma = 1$), the price is set at the monopoly level; if transaction costs are so high that bargaining breaks down ($\gamma = 0$) we are back to the disagreement point.[9] Total flow profit is $q_1 + \gamma q_2$, and must be divided between the two firms. Since the holder of the first patent can block the use of the second invention and can market any good of quality $q \leqslant q_1$, the disagreement point in the bargaining process is $\{q_1,0\}$, where the numbers in brackets denote the profit of firm 1 and 2, respectively. Assuming a fifty-fifty split of the surplus, post-innovation profits before the first patent expires are $\{q_1 + \frac{1}{2}\gamma q_2, \frac{1}{2}\gamma q_2\}$.[10] When the first patent expires, the second innovation no longer infringes. The first innovation becomes public and quality q_1 is marketed at a price equal to marginal costs, i.e. $p_1 = 0$; consequently, the equilibrium price for the second innovation is $p_2 = q_2$.

If $q_2 \geqslant K$, the second innovation is patentable and does not infringe, so the two firms compete in the product market. Assuming Bertrand competition, a limit pricing equilibrium obtains with $p_1 = 0$ and $p_2 = q_2$. In both cases, firm 1's unit profit falls to zero, while firm 2's is equal to q_2. Note the profit erosion associated with product market competition and the corresponding increase in consumers' surplus.

2.5. Social optimum

As a benchmark for future reference, we now characterize the first-best social optimum. We assume that the policymaker controls R&D investment at each stage and decides whether there should be investment in the first period after observing q_1, and in the second period after observing q_2.

Clearly, there must exist cutoff values, x_1 and x_2, such that it is optimal to invest

[9] Heller and Eisenberg (1998) argue forcefully that transaction costs may be far from negligible. They also hold that inefficiencies in the bargaining outcome make the use of leading breadth less attractive. Introducing the parameter γ into our model allows us to assess this contention.

[10] All our results are automatically extended to the case where firm 1 obtains a share α of the bargaining surplus, and firm 2 obtains the remaining share $1 - \alpha$, with $0 < \alpha < 1$.

in the first period if $q_1 \geqslant x_1$ and in the second period if $q_2 \geqslant x_2$. Social welfare can then be written as

$$W = \int_{x_1}^{\bar{q}_1} \left[(2q_1 - c_1) + \int_{x_2}^{\bar{q}_2} (q_2 - c_2) f_2(q_2)\, \mathrm{d}q_2 \right] f_1(q_1)\, \mathrm{d}q_1, \tag{1}$$

where $(2q_1 - c_1)$ is the surplus associated with the first innovation and $(q_2 - c_2)$ is that of the second. Social welfare is the expected value of these surpluses.

The first-best policy can be derived heuristically as follows. The second-period cutoff is obtained by setting the social returns from the second innovation equal to the second-period R&D cost c_2:

$$\hat{x}_2 = c_2. \tag{2}$$

Given the optimal second-period cutoff, the surplus created by the first innovation is

$$2q_1 + \int_{c_2}^{\bar{q}_2} (q_2 - c_2) f_2(q_2)\, \mathrm{d}q_2. \tag{3}$$

The first term is the stand-alone value of the first innovation, i.e. q_1 for two periods; the second term is the option value of the second innovation, i.e. the expected return from the second innovation given the optimal second-period cutoff. Equating (3) and c_1 gives the socially optimal first-period cutoff:[11]

$$\hat{x}_1 = \max \left[\tfrac{1}{2} c_1 - \tfrac{1}{2} \int_{c_2}^{\bar{q}_2} (q_2 - c_2) f_2(q_2)\, \mathrm{d}q_2, 0 \right]. \tag{4}$$

The first-best social optimum $(\hat{x}_1, \hat{x}_2)$ and the social indifference curves are presented in Fig. 1. The $W_{x_2} = 0$ locus is the straight line $x_2 = c_2$; the locus $W_{x_1} = 0$ is increasing for $x_2 > c_2$ and decreasing for $x_2 < c_2$.

[11] Obviously, the same results can be obtained formally by maximizing W with respect to x_1 and x_2. Since

$$W_{x_2} = -[1 - F_1(x_1)](x_2 - c_2) f_2(x_2)$$

and

$$W_{x_1} = -\left[(2x_1 - c_1) + \int_{x_2}^{\bar{q}_2} (q_2 - c_2) f_2(q_2)\, \mathrm{d}q_2 \right] f_1(x_1),$$

the first-order conditions $W_{x_2} = 0$ and $W_{x_1} = 0$ imply Eqs. (2) and (4). It can easily be verified that the second-order conditions hold.

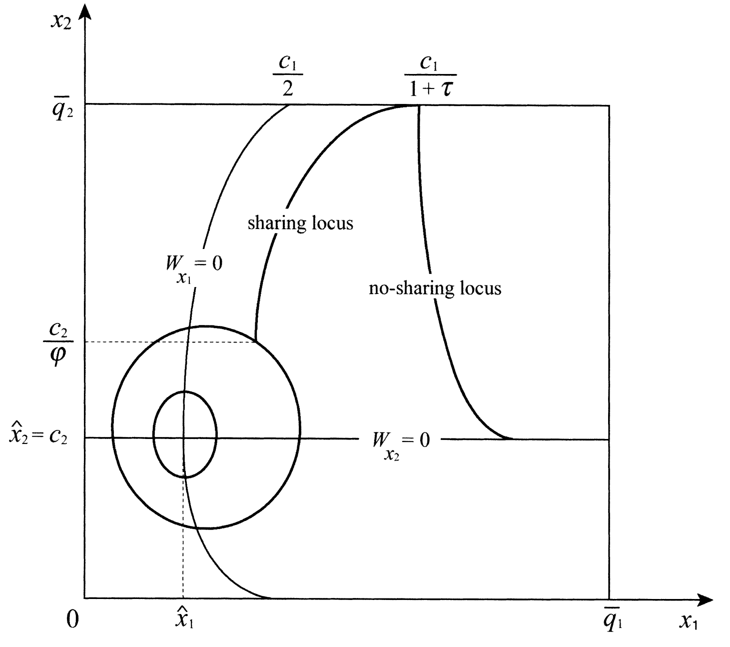

Fig. 1. First-best optimum and market equilibrium in the space of first- and second-period cutoffs. The case in which complete leading breadth is optimal is displayed.

3. Optimal patent policy

In the remainder of the paper we assume that the policymaker does not control x_1 and x_2 directly, but can only choose the patent policy variables, H and K.

3.1. Market equilibrium

As we have ruled out static deadweight losses, social welfare is given by Eq. (1), with the market equilibrium cutoffs x_1^* and x_2^* replacing the socially optimal cutoffs $\hat{x}_1$ and $\hat{x}_2$. Our solution concept is subgame perfect equilibrium, so to derive the market equilibrium for arbitrary values of the policy variables H and K we start from firm 2's decision problem.

Conditional on the occurrence of the first innovation, and having observed q_2, firm 2 must decide whether or not to invest in R&D. Its revenue (profit, gross of R&D costs) is

$$\pi_2 + c_2 = \begin{cases} 0, & \text{for } q_2 < H, \\ \frac{1}{2}\tau\gamma q_2 + (1-\tau)q_2 = \varphi q_2, & \text{for } H \leqslant q_2 < K, \\ q_2, & \text{for } q_2 \geqslant K, \end{cases} \tag{5}$$

where $\varphi \equiv 1 - \tau(1 - \frac{1}{2}\gamma) < 1$. Since π_2 is increasing in q_2, there exists a cutoff value, x_2, such that firm 2 invests if and only if $q_2 \geq x_2$; x_2 is the minimum value of q_2 for which $\pi_2(q_2) \geq 0$. Note that π_2 jumps up at $q_2 = H$ and $q_2 = K$, and increases smoothly for $H < q_2 < K$ and for $q_2 > K$. Thus, we have four cases (see Fig. 2):

$$x_2^* = \begin{cases} H, & \text{for } K \geq H \geq c_2/\varphi, \\ c_2/\varphi, & \text{for } K \geq c_2/\varphi > H, \\ K, & \text{for } c_2/\varphi > K > c_2, \\ c_2, & \text{for } c_2 \geq K. \end{cases} \tag{6}$$

When $c_2 \geq K \geq H$ there is no effective forward protection; in this case, the second-period cutoff is at the socially optimal level, because with no forward protection the second innovator appropriates the entire social return. As the novelty requirement increases, it provides forward protection through a blocking effect: firm 2 will not invest to obtain unpatentable improvements, and this increases the expected length of firm 1's monopoly. However, leading breadth also has a blocking effect: because firm 2 must share the profit accruing from an infringing improvement with firm 1, those infringing improvements whose size is less than $1/\varphi$ times the R&D cost are blocked. This explains why the second-period cutoff depends on K as well as on H.

Working back to the first period, we now determine firm 1's cutoff. Firm 1 observes q_1, knows the distribution function $F_2(q_2)$, and anticipates firm 2's equilibrium investment policy, x_2^*. If $x_2^* \geq K$, there is never infringement. Firm 1's expected profit is then

$$\pi_1 = q_1[1 + \tau F_2(x_2^*)] - c_1. \tag{7}$$

In other words, firm 1's discounted total profit is its first-period profit, plus second-period profit for the duration of the patent if firm 2 does not invest (since firm 2 does not invest if and only if $q_2 < x_2^*$, this event has probability $F_2(x_2^*)$), minus the R&D cost.

If $x_2^* < K$, firm 1's expected profit becomes

$$\begin{aligned} \pi_1 &= q_1 + \tau F_2(x_2^*)q_1 + \tau \int_{x_2^*}^{K} (q_1 + \tfrac{1}{2}\gamma q_2) f_2(q_2)\, \mathrm{d}q_2 - c_1 \\ &= q_1[1 + \tau F_2(K)] + \tfrac{1}{2}\tau\gamma \int_{x_2^*}^{K} q_2 f_2(q_2)\, \mathrm{d}q_2 - c_1. \end{aligned} \tag{8}$$

The additional term on the right-hand-side of Eq. (8) is the profit that firm 1 collects if firm 2 invests but the second innovation infringes (if firm 2 invests and the second innovation does not infringe, i.e. $q_2 \geq K$, firm 1's second-period profit

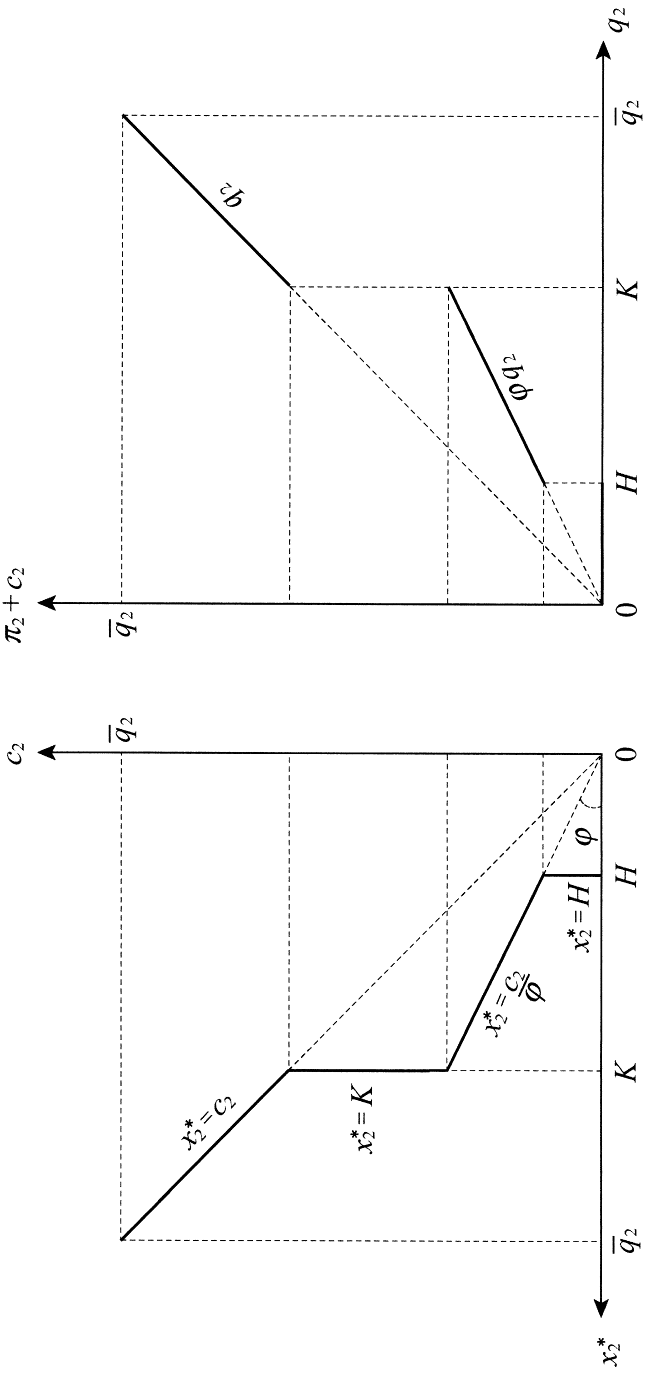

Fig. 2. Second-period gross profit and cutoff.

is nil). This term captures the sharing effect of leading breadth, whereby the intertemporal externality due to the cumulative nature of innovation is partially internalized. The first-period cutoff is then

$$x_1^* = \frac{c_1}{1+\tau F_2(x_2^*)}, \quad \text{for } x_2^* \geqslant K, \tag{9a}$$

$$x_1^* = \max\left[\frac{c_1 - \frac{1}{2}\tau\gamma \int_{x_2^*}^{K} q_2 f_2(q_2)\,\mathrm{d}q_2}{1+\tau F_2(K)}, 0\right], \quad \text{for } x_2^* < K. \tag{9b}$$

Eqs. (6) and (9a,b) fully describe the market equilibrium of the model.

The market equilibrium generally displays underinvestment in R&D.[12] To see this, note that $\hat{x}_2 \leqslant x_2^*$ follows directly by inspection of Eq. (6), with a strict inequality if $K > c_2$. With regard to the first innovation, if $x_2^* \geqslant K$ then the private returns are $q_1[1+\tau F_2(x_2^*)]$ and are lower than the stand-alone social returns from the first innovation $2q_1$; if $x_2^* < K$, using Eqs. (3) and (8) we find that the difference between the social and the private return is

$$q_1[1-\tau F_2(K)] + \int_{c_2}^{x_2^*} (q_2 - c_2) f_2(q_2)\,\mathrm{d}q_2 + \int_{x_2^*}^{K} [q_2(1-\tfrac{1}{2}\tau\gamma) - c_2] f_2(q_2)\,\mathrm{d}q_2 + \int_{K}^{\bar{q}_2} (q_2 - c_2) f_2(q_2)\,\mathrm{d}q_2.$$

All the terms in this expression are always non-negative, and they cannot all go to zero simultaneously. This implies $\hat{x}_1 \leqslant x_1^*$, with a strict inequality if $x_1^* > 0$.

The intuition is that the second innovator can fully appropriate the social returns from the second innovation only if there is no forward protection (otherwise she would have to share the profit with the first innovator, or else part of the social return would spill over in the form of consumers' surplus); but in this case (no forward protection) the first patentee cannot appropriate all of the social value of his innovation. Because the market equilibrium has underinvestment at both stages, the first-best optimum cannot be achieved (unless $c_1 = 0$, in which case the first-best social optimum is obtained for $H = K = 0$). Accordingly we search for the second-best policy.

[12] Similar underinvestment results are found, among others, by Green and Scotchmer (1995), Chang (1995), and O'Donoghue (1998).

3.2. Second-best patent policy

To begin with, note that it is generally desirable to provide some forward protection.

Lemma 1. (positive forward protection) *If* $c_1 > 0$, *a policy of no forward protection* ($H \leqslant K \leqslant c_2$) *is never optimal*.

Proof. Suppose there is no forward protection, and specifically $H \leqslant K = c_2$. From Eqs. (6) and (9a) we then have $x_2^* = c_2$ and $x_1^* = c_1/[1 + \tau F_2(c_2)]$. If K is now increased slightly while leaving H unchanged, the cutoffs become $x_2^* = K$ and $x_1^* = c_1/[1 + \tau F_B(K)]$. To evaluate the effect of this manoeuvre on social welfare we compute

$$\frac{\mathrm{d}W}{\mathrm{d}K} = W_{x_2^*} - W_{x_1^*}\frac{\tau c_1 f_2(c_2)}{[1 + \tau F_2(c_2)]^2}.$$

When $x_2^* = c_2$ and $x_1^* = c_1/[1 + \tau F_2(c_2)]$ we have $W_{x_2^*} = 0$ and $W_{x_1^*} < 0$. This implies that the foregoing derivative must be positive, i.e. the manoeuvre described above must increase social welfare. □

Intuitively, strengthening forward protection will increase investment in the first innovation and reduce it in the second. With no forward protection (and provided that $c_1 > 0$) there is underinvestment in the first innovation, but the second-period cutoff is at its socially optimal level, c_2. Hence, the first-period gain from strengthening forward protection is first order, while the second-period loss is second order.

Having established that some forward protection is desirable, we now ask how it should be provided: via a positive novelty requirement, via a positive leading breadth, or both? The main result of this section is that in the basic model there is no need for a novelty requirement. That is, although there is some latitude in the choice of the policy variables, setting the novelty requirement equal to zero does not prevent the policymaker from achieving the second-best optimum.

Proposition 1. (no novelty requirement) *In the basic model, optimal patent policy does not require a positive novelty requirement*.

Proof. We consider two different cases, $K \leqslant c_2/\varphi$ and $K > c_2/\varphi$. From Eq. (6) it is clear that when $K \leqslant c_2/\varphi$ the equilibrium cutoffs are independent of the novelty requirement, so the policymaker can achieve any feasible value of x_2^* by using the leading breadth alone, setting $H = 0$ (in fact, any $H \in [0,K]$ will do).

When instead $K > c_2/\varphi$, the equilibrium cutoffs may depend on H. However, Eq. (9b) tells us that for any value of K the first-period cutoff is strictly increasing in the second-period cutoff. Since there is underinvestment in R&D, the policy objective is to minimize the cutoffs. This means that the sole efficient policy is to set x_2^*, and hence x_1^*, as low as possible, i.e. $x_2^* = c_2/\varphi$. From Eq. (6) it is clear that in order to achieve $x_2^* = c_2/\varphi$ the policymaker can set the novelty requirement equal to zero (although again there is some latitude in the choice of the policy variables as any policy with $H \in [0, c_2/\varphi]$ will lead to the same equilibrium outcome). □

Proposition 1 can be explained as follows. When $K \leqslant c_2/\varphi$, firm 2 invests only if there is no infringement. This means that leading breadth has a blocking but no-sharing effect. In this case, novelty requirement and leading breadth are perfect substitutes, as both provide forward protection exclusively by blocking second-period improvements. However, when $K > c_2/\varphi$, leading breadth has a sharing effect, although it inevitably blocks some improvements. Since the first innovator actually benefits from the arrival of infringing second-period innovations, for any given value of K it is clearly inefficient to use the novelty requirement to block any further improvements.

This argument is robust to the existence of deadweight losses associated with monopoly pricing, which would arise with elastic demand.[13] The same conclusion holds if ex ante agreements (i.e., agreement reached before firm 2's investment is sunk) are feasible, as in Green and Scotchmer (1995) and Scotchmer (1996).[14]

3.3. Comparative statics

In the remainder of this section we provide a fuller characterization of the optimal patent policy in our basic model. With no loss of generality, we set $H = 0$. It is useful to note the following lemma.

[13] The existence of such deadweight losses implies that social welfare no longer depends only on the first- and second-period cutoffs but also on whether patentees compete or collude, and is greater with competition. However, it remains true that the novelty requirement and leading breadth are perfect substitutes when forward protection is provided exclusively by blocking, and that it is inefficient to use the novelty requirement to block some further innovation under complete leading breadth (with elastic demand, consumers' surplus generally increases when the second innovation arrives even if there is monopoly pricing). The same is true of the case where inefficiency in bargaining takes the form of a waste of real resources rather than pricing below the monopoly level.

[14] The argument for this case is different. With ex ante agreements, the blocking effect of the novelty requirement disappears: efficient development of the second innovation is guaranteed, and the second-period cutoff is always c_2. However, the first innovator cannot capture the entire return, so the social problem becomes one of transferring as much as possible of the total profit from the two inventions to the original innovator. This requires a complete leading breadth, as argued by Green and Scotchmer (1995), but, rather surprisingly, no novelty requirement. A positive novelty requirement would give firm 2 a credible threat not to invest even when second-period investment is jointly optimal, increasing its bargaining position in the ex ante agreement.

Lemma 2. (sharing implies complete leading breadth) *The optimal policy involves either* $K \leq c_2/\varphi$ *or* $K = \bar{q}_2$.

Proof. When $K > c_2/\varphi$, x_2^* is independent of K while the first-period cutoff depends on x_2^* and K, as is clear from Eq. (9b). Since there is always underinvestment in the first innovation, K must be chosen so as to minimize the first-period cutoff x_1^* for any given x_2^*. This clearly requires that K be set as high as possible, i.e. $K = \bar{q}_2$, so that all second-period innovations infringe. □

The intuition is that when $K > c_2/\varphi$, the second-period cutoff is independent of leading breadth, so raising K does not impede second-period innovations. Complete leading breadth shifts as much as possible of the profit from the second innovation to the first patentee, thus stimulating first-period innovation at zero social cost.

Lemmas 1 and 2 imply that the policy problem reduces to choosing between $K = \bar{q}_2$ (*sharing with complete leading breadth*) and the optimal value of K in the interval $(c_2, c_2/\varphi)$ (*blocking with incomplete leading breadth*). We then address two issues: under what circumstances is complete or incomplete leading breadth optimal? And when incomplete leading breadth is optimal, how does the optimal value of K change with the model's parameters?

The policymaker's problem is depicted in Fig. 1. For $K = \bar{q}_2$, Eq. (9b) defines a strictly increasing locus in the (x_1, x_2) space. Along this locus, firm 2 may invest to obtain infringing improvements, and so we shall call this the "sharing locus." Likewise, Eq. (9a) defines a strictly decreasing locus in the (x_1, x_2) space. Here there is never infringement, hence no profit sharing, so we call it "no-sharing locus." (The underinvestment result means that both the no-sharing and the sharing loci must lie above and to the right of the social optimum in Fig. 1.) The first issue we address is to determine under what conditions the social optimum lies on the sharing or the no-sharing locus.

To gain more insight into the social trade-off, note that with profit sharing second-period improvements are not stifled, the intertemporal externality due to the cumulative nature of innovation is at least partially internalized, and there is no profit erosion due to competition between the patentees. These effects explain why the sharing locus lies to the left of the no-sharing locus. However, the minimum second-period cutoff along the sharing locus is c_2/φ while along the no-sharing locus it is c_2. Since $\varphi < 1$, profit sharing is attainable only if society is prepared to give up more valuable second-period improvements (this follows from the blocking effect of leading breadth). Sharing may even be unattainable if $c_2 > \varphi\bar{q}_2$, in which case incomplete leading breadth is necessarily optimal.

The next proposition shows how the optimality of complete or incomplete leading breadth depends on the costs of the innovations. As noted, the social optimum must lie on the no-sharing locus for $c_2 > \varphi\bar{q}_2$. When $c_2 \leq \varphi\bar{q}_2$, we have:

Proposition 2. (sufficient conditions for complete or incomplete leading breadth)

(*i*) *When* c_1 *is sufficiently small, the optimal leading breadth is incomplete*: $K \in (c_2, c_2/\varphi)$.

(*ii*) *When* c_2 *is close to* 0 *or* c_1 *is close to* $2\bar{q}_1$, *complete leading breadth is optimal*: $K = \bar{q}_2$.

Proof.

(i) When $c_1 = 0$, a policy of no forward protection replicates the first-best social optimum (setting $K = H = 0$, the equilibrium cutoffs are $x_1^* = 0$ and $x_2^* = c_2$, coinciding with the socially optimal). By continuity, weak forward protection (and hence incomplete leading breadth) continues to be desirable for c_1 close to 0.

(ii) When $c_2 = 0$, leading breadth has no blocking effect and sharing is therefore strictly better than blocking (unless $c_1 = 0$, in which case both firms always invest provided that $H = 0$, independently of the leading breadth). The optimality of complete leading breadth for low values of c_2 then follows by continuity.

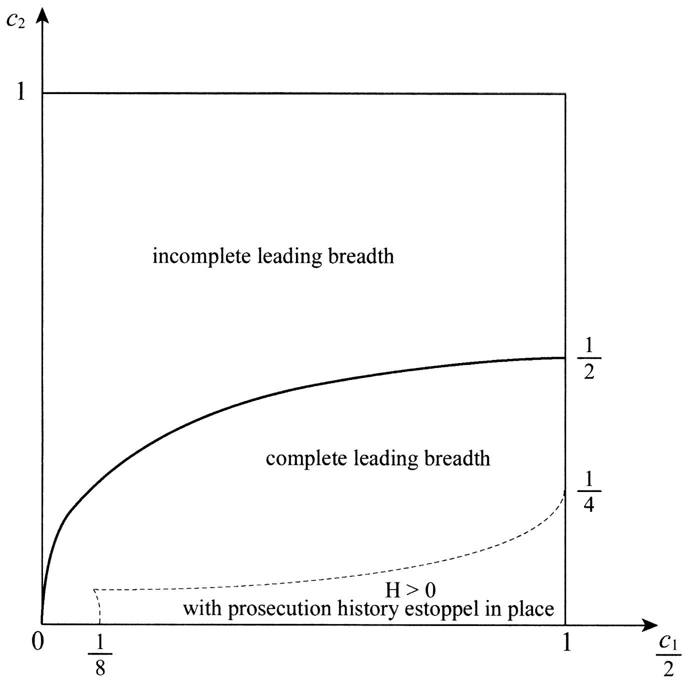

Fig. 3. Optimal policy in the uniform distribution case, $\tau = 1$ and $\gamma = 1$.

When c_1 is close to $2\bar{q}_1$, if leading breadth is incomplete so that $K < c_2/\varphi$, firm 1 will almost never invest, because it garners no rent from the second innovation; social welfare would thus be close to 0. In contrast, provided that $c_2 \leqslant \varphi\bar{q}_2$, with $K = \bar{q}_2$ social welfare does not fall to zero even if c_1 approaches $2\bar{q}_1$, because the prospect of sharing the profit from the second innovation provides firm 1 with the incentive to invest. □

Intuitively, if c_1 is small, there is little need to protect the first innovator since first-period profit alone provides the incentive to invest. Complete leading breadth is attractive when the first innovation is costly, so that first-period profit fails to recoup R&D spending. It is also attractive when the cost of the second innovation is low, because this means that the option value of the second innovation is high and the intertemporal externality accordingly strong, and because when c_2 is large the blocking effect of leading breadth is socially costly. Proposition 2 means that the optimal policy will switch to complete leading breadth as c_2 decreases, or as c_1 increases (provided that $c_2 < \varphi\bar{q}_2$). It can be shown that under mild regularity conditions there can be no reswitching.[15] Fig. 3 illustrates the set of parameter values supporting complete or incomplete leading breadth for the case of uniform and identical probability distributions over the unit interval [0,1], $\tau = 1$ and $\gamma = 1$.[16]

Now suppose that the optimal leading breadth is incomplete. How does the optimal value of K change with the costs of the innovations?

From Eqs. (6) and (9a) we get $x_2^* = K$ and $x_1^* = c_1/[1 + \tau F_2(K)]$. Substituting into Eq. (1) and differentiating we get

$$\frac{\mathrm{d}W}{\mathrm{d}K} = \frac{\tau c_1}{[1 + \tau F_2(K)]^2}\left[\frac{2c_1}{[1 + \tau F_2(K)]} - c_1 + \int_K^{\bar{q}_2} (q_2 - c_2) f_2(q_2)\,\mathrm{d}q_2\right] f_1\left(\frac{c_1}{1 + \tau F_2(K)}\right) f_2(K) - \int_{c_1/[1+\tau F_2(K)]}^{\bar{q}_1} (K - c_2) f_2(K) f_1(q_1)\,\mathrm{d}q_1.$$

[15] Details are available from the authors upon request.

[16] Relative to Fig. 3, the existence of deadweight losses associated with monopoly pricing (with elastic demand) would make incomplete leading breadth more attractive. In this case, two new effects come into play. First, Lemma 2 need no longer hold. Now the optimal choice of K has to balance the gains accruing from faster technological progress against the welfare loss associated with monopoly in the use of the new technology — a familiar trade-off in the theory of patents. To the extent that the optimal value of K is lower than $\bar{q}_2$, the relevant sharing locus shifts to the right. Second, the size of the deadweight loss depends on whether patentees compete or collude; it is larger with collusion. Both effects tend to favor blocking.

Table 1
Optimal leading breadth with uniform distributions, $\tau = 1$ and $\gamma = 1$

c_2	c_1					
	0.2	0.6	0.9	1.2	1.5	1.8
0.15	0.156	1	1	1	1	1
0.30	0.305	1	1	1	1	1
0.45	0.452	0.571	0.667	0.765	1	1
0.60	0.601	0.667	0.732	0.806	0.882	0.954
0.75	0.750	0.780	0.815	0.861	0.913	0.966
0.90	0.900	0.908	0.920	0.937	0.959	0.983

The first term is the social gain from increasing K associated with the increase in first-period R&D investment; the second term is the social loss from the fall in second-period investment. The optimal leading breadth is found by balancing these two effects. In the case of uniform distributions, $\gamma = \tau = 1$, the optimal value of K can be calculated numerically (see Table 1).

Table 1 shows that the optimal leading breadth is increasing in c_1 but is non-monotonic in c_2. While an increase in c_2 tends to make complete leading breadth relatively less attractive, when incomplete leading breadth is optimal, K tends to increase with c_2. The reason is that the second-period cutoff x_2^* should tend to increase with c_2, and with incomplete leading breadth we have $x_2^* = K$.[17]

Finally, note that as γ decreases, i.e. as bargaining becomes less efficient, the sharing locus associated with $K = \overline{q}_2$ shifts to the right, while the blocking locus is unchanged. This means that high transaction costs are socially costly, and tend to make weak forward protection preferable (although it remains optimal to provide forward protection through leading breadth only). With regard to patent length, both the sharing and the no-sharing loci shift to the right as τ decreases. The effect on the optimal strength of forward protection is ambiguous.

4. Collusion

In the model of the previous section, one important effect of complete leading breadth is to prevent destructive competition between the patentees. One may argue that if collusion is permitted whether or not there is an infringement, this function is eliminated and the novelty requirement may come to play a more active role. Moreover, if collusion is permitted there may be an inefficient incentive to

[17] Given Proposition 1, our characterization of the optimal policy may be compared to Chang's, which focuses on leading breadth only. Although our results are broadly consistent with his (in particular, the non-monotonicity of optimal breadth in c_2 recalls Chang's result that K should be non-monotonic in q_1), the set of exogenous parameters in our model (i.e., c_1 and c_2) differs from his, so it is not possible to draw precise analogies. For example, in the model of this section, K is independent of q_1.

invest in trifling improvements that can serve as a bargaining weapon to appropriate some of the original innovator's profit. It has long been argued (Beck, 1976) that patent policy should block such nuisance improvements, for which the novelty requirement may seem appropriate. In this section we address those concerns. We show that permitting collusion even in the absence of infringement makes incomplete leading breadth more attractive, but does not alter the result that optimal policy does not require a positive novelty requirement.

Suppose, then, that collusive agreements are permitted even if the second innovation does not infringe. Nothing changes when $q_2 < K$, so Eq. (9b) and Lemma 2 continue to hold. Let us therefore focus on the case in which $q_2 \geqslant K$. Now the two patentees can collude even if the second innovation does not infringe. In this case, the disagreement point is $\{0, q_2\}$, since firm 2 can use its innovation without the consent of firm 1. The surplus to be divided is q_1, and assuming again a fifty-fifty split, the outcome is $\{\frac{1}{2}\gamma q_1, q_2 + \frac{1}{2}\gamma q_1\}$.

Hence, with collusion, firm 2's profit if the second innovation does not infringe is $\pi_2 = q_2 + \frac{1}{2}\tau\gamma q_1 - c_2$, and the second-period cutoff is

$$x_2^* = \begin{cases} K, & \text{for } c_2 \leqslant K + \frac{1}{2}\tau\gamma q_1, \\ c_2 - \frac{1}{2}\tau\gamma q_1, & \text{for } c_2 \geqslant K + \frac{1}{2}\tau\gamma q_1. \end{cases} \tag{10}$$

Note that in the absence of forward protection x_2^* now falls below c_2; in other words, when collusion is permitted there may be overinvestment in the second innovation.[18] Small improvements are now strategically valuable; they serve as a bargaining threat to steal some of the first innovator's profit. Reasoning as in the proof of Lemma 1, we can easily show a fortiori that some forward protection should be provided.

Working back to the first period, firm 1's expected profit is now

$$\pi_1 = q_1 + \tau F_2(x_2^*)q_1 + \tfrac{1}{2}\tau\gamma q_1[1 - F_2(x_2^*)] - c_1. \tag{11}$$

The first-period cutoff is therefore

$$x_1^* = \frac{c_1}{1 + \tau[\frac{1}{2}\gamma + (1 - \frac{1}{2}\gamma)F_2(x_2^*)]}. \tag{12}$$

Comparing Eqs. (12) and (9a), one sees immediately that with collusion the no-sharing locus shifts to the left but still lies to the right of the sharing locus: even if aggregate profits are now the same with and without infringement, the disagreement point when the second innovation infringes is more favorable to firm 1.

The fact that the no-sharing locus shifts leftward has two corollaries. First, in

[18] Note also that the second-period cutoff is no longer independent of the size of the first innovation q_1, because when collusion is allowed, the follow-on firm may obtain a share of the profit from the original innovation.

our model permitting collusion is always desirable.[19] Second, since the sharing locus is unaffected, incomplete leading breadth is relatively more attractive under collusion. Nevertheless, Proposition 1 still holds.

5. The role of the novelty requirement

We now analyze some model extensions that create a more active role for the novelty requirement. First, we identify one case where it has no blocking effect, namely, that of repeated innovations. In this case, the optimal patent policy does not require a positive effective leading breadth. Next, we analyze the possibility that both instruments must be used simultaneously. To save on notation, hereafter we set $\gamma = 1$ and $\tau = 1$.

5.1. Repeated innovations

Suppose that the first innovator can compete for the second innovation on an equal footing with the outsider, as in Denicolò (2000). In this variant, imposing a strong novelty requirement ($H = K = \bar{q}_2$) does not block the second innovation. If the latter is unpatentable and infringes, then only the first patent-holder can lawfully use it, so he alone has an incentive to invest in the improvement. Thus, making the second innovation unpatentable isolates the first innovator from R&D competition, allowing him to capture the entire option value of the second innovation. The private returns from the innovations would then equal the social returns, so by setting $H = K = \bar{q}_2$ the policymaker achieves the first-best social optimum.

5.2. Investment under uncertainty

Suppose that firms invest in R&D before the uncertainty over the value of innovations is resolved. Firms' investment decision then becomes a 0–1 choice, and the analysis of optimal patent policy is considerably simplified. The first-best social optimum may require firms not to invest in the first innovation, to invest in the first but not the second, or to invest in both. Since the private return from innovations cannot exceed the social return, implementing the first-best optimum is straightforward in the first two cases.

When the social optimum requires that both innovations must be achieved, however, things are more complex. In this case, the expected net social return from

[19] This is unsurprising in view of our assumption that monopoly entails no deadweight loss. Even if overinvestment in the second innovation may occur, it can be eliminated at zero social cost through the blocking effect of leading breadth and therefore only the beneficial effect of collusion remains (namely, there is no profit erosion).

the second innovation is positive, and with no forward protection firm 2 invests. If firm 1 invests with $K = H = 0$, the optimal policy involves no forward protection and leads to the first-best.[20] However, if firm 1 would not invest with $K = H = 0$, some rents must be transferred to the first innovator: that is, positive forward protection is needed.

We know that raising K leads to profit sharing. When firm 2 invests before the uncertainty over q_2 is resolved, however, raising H also allows the first innovator to capture some of the rents from the second innovation, because firm 2 may now generate unpatentable improvements. Since these are patent infringements (recall that $K \geqslant H$), only firm 1 can use them. As H increases (provided that firm 2 continues to invest) firm 1 can free-ride on more second-period improvements. This means that the novelty requirement has now both a blocking and a sharing effect.

Starting from $H = K = 0$, raising K is a more efficient way of transferring rents to the first innovator than raising H. The reason is that as H increases (with K unchanged), rents are transferred from the second to the first innovator on a one-to-one basis because aggregate profits do not change (for infringing innovations, the equilibrium price is $q_1 + q_2$ whether or not the second innovation is patentable). By contrast, raising K allows firms to collude in a wider set of circumstances and this leads to greater aggregate profits, alleviating the problem of underinvestment. As a consequence, only leading breadth should be used if it suffices to induce firm 1 to invest.[21] However, if firm 1's net expected profit is negative even with complete leading breadth and $H = 0$ (and the second inventor's expected profits stay positive), raising H may become socially desirable.[22]

To summarize, when firms invest in R&D before uncertainty is resolved, leading breadth by itself suffices when optimal forward protection is relatively weak, but the novelty requirement (with complete leading breadth) becomes necessary when the optimal level of forward protection is sufficiently high.

[20] Recall that we assume a complete lagging breadth, which protects the first innovator against imitators. We recognize that in practice with $H = K = 0$ the fact that even very small improvements are patentable and do not infringe may allow imitators to camouflage, but in the model we assume that the Patent Office and the courts can distinguish imitation from genuine — albeit trivial — improvements.

[21] Note that Lemma 2 fails when firms invest before uncertainty is resolved, and so complete leading breadth is not necessarily optimal.

[22] The first-best optimum may not be attainable, however. If firm 2's net expected profit with $K = \bar{q}_2$ and $H = 0$ is negative, incomplete leading breadth is necessary for investment in the second innovation to be profitable. In this case some profit erosion is inevitable, and the private returns from the innovations will fall short of the social returns. This means that it may be impossible to induce both firms to invest even if the social value of both innovations is positive.

5.3. Prosecution history estoppel

So far we have implicitly assumed that both innovations have complete *lagging breadth*.[23] Let us now explore the hypothesis that the second patent's lagging breadth is incomplete.[24]

In particular, we focus on the legal doctrine of ''prosecution history estoppel'' or ''file wrapper estoppel,'' which links the novelty requirement H with the lagging breadth of the second patent. Suppose that a second-period innovation of size q_2 effectively makes it possible to market any good of quality $q \leqslant q_1 + q_2$. Suppose also that the improvement is not obvious, i.e. $q_2 > H$. Nevertheless, all the techniques in the interval $[q_1, q_1 + H]$ may be regarded as obvious improvements on the first innovation: this, in fact, is the rationale for setting the novelty requirement equal to H. Thus, the courts have systematically ruled that the second innovator cannot claim the techniques in the interval $[q_1, q_1 + H]$ even if she is the true and original inventor — for otherwise the patent would be invalid for lack of novelty — and must make them available to the public.[25] (In the present context, $K \geqslant H$ means that all these techniques infringe on the first patent so that only the first innovator can actually exploit them.)

Let us proceed to an economic analysis of prosecution history estoppel. How does the optimal policy mix change when prosecution history estoppel is in being? In particular, can the novelty requirement now play a more active role? We can show that social welfare increases when prosecution history estoppel is provided for.

Prosecution history estoppel means that the range of lagging breadth cannot exceed $q_2 - H$. We assume that it is equal to $q_2 - H$; it will become clear in the sequel that this entails no loss of generality. Then, if the second innovation

---

[23] A patent disclosing the best new technology may also make available to the public a whole set of dominated techniques that are still superior to the prior art; the lagging breadth specifies which of these are covered by the patent. See O'Donoghue et al. (1998) for a detailed discussion of lagging breadth.

[24] In the model, it is always optimal to protect the first innovation with a complete lagging breadth, since reducing it would stifle first-period innovation with no countervailing social benefit.

[25] See *Autogiro co. of America v. United States*, 384 F.2d 391, 155 U.S.P.Q. 697 (Ct. Cl. 1967). The court ruled that

> *a claim may not include within its range anything that would violate limitations expressed before the Patent Office. Thus a patent that has been severely limited to avoid the prior art will only have a small range between it and the point beyond which it violates file wrapper estoppel.*

The court also ruled that ''the doctrine of equivalents [stating that minimal differences between the accused product or method and the claims of a patent do not prevent infringement] is subservient to file wrapper estoppel.'' Also related is the recapture rule, which prevents an applicant from seeking claims on reissue of a patent that were abandoned in the original application: see, for instance, *Ball Corp. v. United States*, 729 F.2d 1429, 221 U.S.P.Q. (BNA) 289 (Fed. Cir. 1984). See generally Merges (1997).

infringes on the original patent, the disagreement point in the bargaining process is $\{q_1 + H, 0\}$, so post-innovation profits are $\{q_1 + \frac{1}{2}(q_2 + H), \frac{1}{2}(q_2 - H)\}$. If the second innovation does not infringe, the limit pricing equilibrium under Bertrand competition is now $p_1 = 0$ and $p_2 = q_2 - H$. Consequently, firm 2's profit if it invests becomes (to save on notation, hereafter we set $\gamma = 1$ and $\tau = 1$; this implies $\varphi = \frac{1}{2}$)

$$\pi_2 = \begin{cases} -c_2, & \text{for } q_2 < H, \\ \frac{1}{2}(q_2 - H) - c_2, & \text{for } H \leqslant q_2 < K, \\ q_2 - H - c_2, & \text{for } q_2 \geqslant K, \end{cases} \tag{13}$$

and the second-period cutoff is

$$x_2^* = \begin{cases} H + 2c_2, & \text{for } \Delta \geqslant 2c_2, \\ K, & \text{for } c_2 \leqslant \Delta \leqslant 2c_2, \\ \min[H + c_2, \bar{q}_2], & \text{for } \Delta \leqslant c_2, \end{cases} \tag{14}$$

where $\Delta \equiv K - H$ is the effective leading breadth.

If $x_2^* \geqslant K$, the equilibrium does not change. If $x_2^* < K$, firm 1's expected profit is

$$\pi_1 = q_1[1 + F_2(K)] + \int_{x_2^*}^{K} \tfrac{1}{2}(q_2 + H) f_2(q_2)\, \mathrm{d}q_2 - c_1. \tag{15}$$

Eq. (9b) is now replaced by

$$x_1^* = \max\left[\frac{c_1 - \int_{H+2c_2}^{K} \frac{1}{2}(q_2 + H) f_2(q_2)\, \mathrm{d}q_2}{1 + F_2(K)}, 0\right]. \tag{16}$$

The underinvestment result and Lemma 1 continue to hold.

The policy space $0 \leqslant H \leqslant K \leqslant \bar{q}_2$ can be divided into two regions, $\Delta > 2c_2$ and $\Delta \leqslant 2c_2$ (see Fig. 4). When $\Delta \leqslant 2c_2$ we have $x_2^* \geqslant K$, so firm 2 invests only if the second innovation does not infringe; thus, there is no profit sharing. The equation of the no-sharing locus is still given by Eq. (9a). Provided that $\Delta \leqslant 2c_2$, all policy combinations of H and K that lead to the same value of x_2^* are equivalent in terms of social welfare. Fig. 4 depicts these inverse L-shaped equivalence curves. Clearly, the policymaker can achieve any feasible value of x_2^* in this region by using just the novelty requirement, i.e. by setting $\Delta = 0$. Less obviously, he can use the leading breadth only, setting $H = 0$ (although this way one cannot achieve $x_2^* > 2c_2$, the optimal policy can lie on the no-sharing locus only if the second-period cutoff satisfies $x_2^* < 2c_2$).

When $\Delta > 2c_2$, some second-period innovations will infringe and there will be profit sharing. Here, the first-period cutoff depends both on x_2^* and on K. From Eq. (14) it follows that $x_2^* = H + 2c_2$ and substituting into Eq. (16) we get

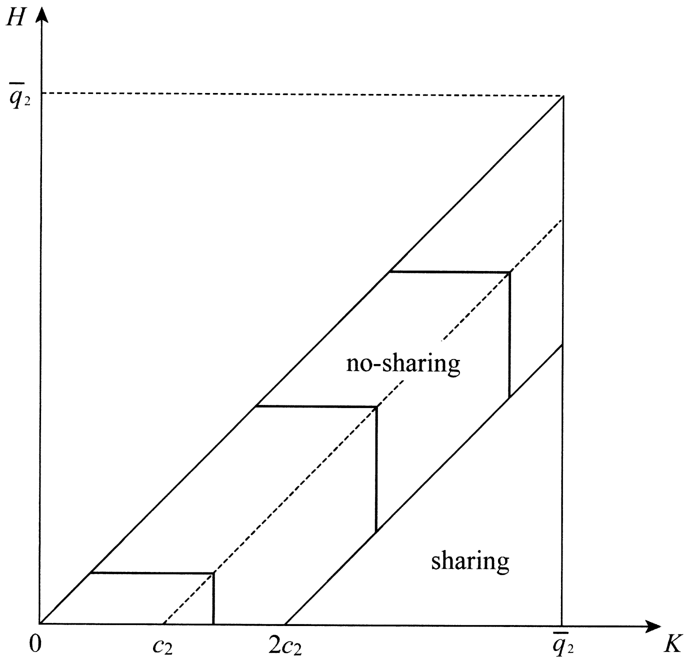

Fig. 4. The policy space under prosecution history estoppel.

$$x_1^* = \max\left[\frac{c_1 - \int_{x_2^*}^{K} (\frac{1}{2}q_2 + \frac{1}{2}x_2^* - c_2)f_2(q_2)\,\mathrm{d}q_2}{1 + F_2(K)}, 0\right]. \tag{17}$$

Clearly, Lemma 2 continues to hold, and therefore $K = \bar{q}_2$,[26] but we can no longer be sure that the sharing locus is always increasing, in that raising H now has two conflicting effects on firm 1's profit, the familiar blocking effect and a sharing effect associated with prosecution history estoppel. More precisely, the blocking

[26] We can now confirm that there is no loss of generality in setting the lagging breadth of the second patent equal to its maximum feasible value $q_2 - H$. To show this, suppose that the breadth is $q_2 - L$, with $L > H$. Then, when $x_2^* < K$ we have $x_2^* = L + 2c_2$ and

$$\pi_1 = q_1[1 + F_2(K)] + \tfrac{1}{2}\int_{x_2^*}^{K} (q_2 + L)f_2(q_2)\,\mathrm{d}q_2 - c_1.$$

Lemma 2 continues to hold, so setting $K = \bar{q}_2$ and substituting $x_2^* = L + 2c_2$ into the above expression we re-obtain Eq. (17), independently of the value of L. If $x_2^* \geqslant K$, it is clear that firm 1's profit is still given by Eq. (7) and therefore Eq. (9a) continues to hold. This means that setting $L > H$ does not change the no-sharing and sharing loci, and therefore there is no loss of generality in setting $L = H$.

effect is that the second period cutoff x_2^* increases with H, so the first innovator gets less rent from the second innovation as the latter occurs less frequently. The sharing effect is that as H increases, the share of the returns from the second innovation accruing to the first innovator increases because the disagreement point in the bargaining process becomes more favorable to him.

In principle, either effect may prevail. It can be shown that the blocking effect must prevail for x_2^* close to $\bar{q}_2$.[27] However, if c_2 is sufficiently low, the sharing locus may bend backward as x_2^* decreases; that is, it will eventually bend backward provided that

$$c_2 < \frac{1 - F_2(2c_2)}{2f_2(2c_2)}.$$

While the conditions that lead to complete and incomplete leading breadth are qualitatively the same as in Proposition 2, we can now show that $H > 0$ may be part of the optimal policy.

Proposition 3. (sufficient condition for positive novelty requirement) *With prosecution history estoppel in place, when* c_2 *is close to* 0 *and* $c_1 > \frac{1}{2}\int_0^{\bar{q}_2} q_2 f_1(q_2)\,\mathrm{d}q_2$, *the optimal novelty requirement H is positive.*

Proof. Arguing as in the proof of Proposition 2 we can show that complete leading breadth is optimal for c_2 sufficiently low. To show that the solution on the sharing locus is interior when $c_2 = 0$ and $c_1 > \frac{1}{2}\int_0^{\bar{q}_2} q_2 f_1(q_2)\,\mathrm{d}q_2$, suppose to the contrary that the solution is at the corner with $H = 0$ and $K = \bar{q}_2$. This implies $x_2^* = 0$ and $x_1^* = \frac{1}{2}c_1 - \frac{1}{4}\int_0^{\bar{q}_2} q_2 f_2(q_2)\,\mathrm{d}q_2$. Beginning from this point, by raising H we can increase x_2^*, and the corresponding fall in x_1^* is $\mathrm{d}x_1^*/\mathrm{d}x_2^* = -\frac{1}{4}$. Social welfare changes as

$$\frac{\mathrm{d}W}{\mathrm{d}x_2^*} = W_{x_2^*} - \frac{1}{4}W_{x_1^*}.$$

Since at $x_2^* = 0$ and $x_1^* = \frac{1}{2}c_1 - \frac{1}{4}\int_0^{\bar{q}_2} q_2 f_2(q_2)\,\mathrm{d}q_2$ we have $W_{x_2^*} = 0$ and $W_{x_1^*} < 0$, the increase in H is welfare-improving, implying that the solution must be interior. □

Fig. 3 displays the region where the optimal policy involves $H > 0$ in the uniform distribution case, with $\tau = \gamma = 1$. The intuition underlying this result is

[27] When $K = \bar{q}_2$, differentiation of Eq. (17) leads to

$$\frac{\mathrm{d}x_1^*}{\mathrm{d}x_2^*} = \frac{1}{2} f_2(x_2^*)(x_2^* - c_2) - \frac{1}{4}[1 - F_2(x_2^*)].$$

For x_2^* close to $\bar{q}_2$, $F_2(x_2^*)$ will be close to 1 and the derivative is positive.

that when c_2 is low, the intertemporal externality is very strong (the option value of the second innovation is large), and even with complete leading breadth it would be desirable to strengthen forward protection. Then, with prosecution history estoppel it becomes desirable to use the novelty requirement because of its sharing effect.

Finally, note that setting $H = 0$ was optimal with complete lagging breadth and is still feasible under our current assumptions. The fact that now it may be optimal to set $H > 0$ means that social welfare is higher with prosecution history estoppel.

6. Conclusion

In this paper we have analyzed the optimal combination of the novelty requirement and leading breadth to provide forward patent protection in a two-stage model of innovation. In our basic model, the novelty requirement and leading breadth are perfect substitutes for one another when the level of forward patent protection is low. As forward protection strengthens, only leading breadth should be used. The reason is that broad leading breadth leads to profit sharing while the novelty requirement operates exclusively through a blocking mechanism, and sharing is preferable to blocking. This result holds regardless of the duration of the patent, transaction costs and deadweight losses, and whether or not collusion between patentees is permitted.

However, we have also identified cases in which the novelty requirement may also form part of the optimal patent policy. With repeated innovations, the optimal policy actually involves no effective leading breadth. When prosecution history estoppel is in place, the two instruments should be used together when optimal forward patent protection is very strong. The same holds when firms invest in R&D before uncertainty over the value of innovations is resolved.[28] In all of these cases, the novelty requirement has a sharing effect (with or without its usual blocking effect).

Our model leads to a definite policy conclusion: the novelty requirement should be used only when strong forward protection is optimal and when leading breadth

[28] O'Donoghue (1998) also finds that a positive novelty requirement can increase forward patent protection even starting from complete leading breadth. O'Donoghue models a long sequence of innovations; each innovation builds on the previous one and opens the way to the subsequent improvement, so the social problem is to determine the level of profit for each innovator, not the division of profits between first- and second-generation innovators. He also assumes that firms choose both the level of R&D investment and the size of innovations that they target. In such a framework, he shows that strengthening the novelty requirement forces firms to target larger innovations, thus increasing forward protection. This is essentially a blocking mechanism, while in our model the novelty requirement becomes desirable only when it has a sharing effect.

is already set at its maximum feasible level.[29] This conclusion, however, rests on many simplifying assumptions that we have made to render the model tractable. More research is required to draw practical implications.

For example, we have ruled out competition in R&D. We have assumed perfect and costless enforcement of patent rights, whereas it is well known that in real life patent enforcement is far from perfect and litigation costs are substantial. Although we have tested the robustness of our conclusions in a number of ways, we have not allowed for the interaction of different effects. Finally, we have restricted our attention to the policy instruments that are available given current patent law, not asking whether more radical reforms of the patent system may be desirable.[30] All of these issues are left for future research.

Acknowledgements

We are grateful to two anonymous referees, Steven Martin, and seminar audiences at Helsinki, Padova, and Milano for helpful comments. Vincenzo Denicolò gratefully acknowledges financial support from MURST.

References

Beck, R., 1976. Patents, property rights, and social welfare: search for a restricted optimum. Southern Economic Journal 43, 1045–1055.

Bessen, J., Maskin, E., 2000. Sequential innovations, patents, and imitation. MIT, mimeo.

Chang, H., 1995. Patent scope, antitrust policy, and cumulative innovation. RAND Journal of Economics 26, 34–57.

Denicolò, V., 2000. Two-stage patent races and patent policy. RAND Journal of Economics 31, 488–501.

Green, J., Scotchmer, S., 1995. On the division of profit in sequential innovation. RAND Journal of Economics 26, 20–33.

[29] However, we have also shown that focusing solely on the novelty requirement has no cost when the level of forward protection is low. This may shed some light on certain controversial provisions of the Semiconductor Chip Protection Act (SCPA) of 1984. Prior to the SCPA, most semiconductor designs failed to meet the novelty requirement applied by the Patent Office or the courts. Since copyrights are inapplicable to semiconductors, and trade secrets are also not well suited because semiconductors can be easily reverse-engineered, the SCPA introduced mask works as a new form of intellectual property. The policy objective was clearly to weaken forward protection in this particular industry: mask works have a shorter duration and much less stringent novelty requirements than patents. As Hunt (1999) notes, ‘‘the SCPA took the unusual step of making mask work immune from the risk of infringing earlier works,’’ a provision that is unique to the SCPA and has been criticized by some commentators, but might be justified on the basis of our findings.

[30] See Hopenhayn et al. (2000) for a mechanism design approach to innovation policy with sequential innovation.

Heller, M.A., Eisenberg, R.S., 1998. Can patents deter innovation? The anticommunist in biomedical research. Science 280, 698–701.

Hopenhayn, H., Llobet, G., Mitchell, M., 2000. Rewarding sequential innovators: prizes, patents and buyouts. Mimeo, University of Rochester.

Hunt, R.M., 1999. Nonobviousness and the incentive to innovate: an economic analysis of intellectual property reform. Federal Reserve Bank of Philadelphia Working Paper, #99-3.

Matutes, C., Regibeau, P., Rockett, K.E., 1996. Optimal patent protection and the diffusion of innovation. RAND Journal of Economics 27, 60–83.

Merges, R.P., 1997. Patent Law and Policy. Michie, Charlottesville.

O'Donoghue, T., 1998. A patentability requirement for sequential innovations. RAND Journal of Economics 29, 654–679.

O'Donoghue, T., Scotchmer, S., Thisse, J.F., 1998. Patent breadth, patent length, and the pace of technological progress. Journal of Economics and Management Strategy 7, 1–32.

O'Donoghue, T., Zweimuller, 1998. Patents in a model of endogenous growth. CEPR Discussion Paper #1951.

Scotchmer, S., 1991. Standing on the shoulders of giants: cumulative research and the patent law. Journal of Economic Perspectives 5, 29–41.

Scotchmer, S., 1996. Protecting early innovators: should second-generations products be patentable? RAND Journal of Economics 27, 322–331.

Scotchmer, S., 1999. Cumulative innovation in theory and practice. University of California, Berkeley, mimeo.

van Dijk, T., 1995. Patent height and competition in product improvements. Journal of Industrial Economics 44, 151–167.