RAND Journal of Economics
Vol. 32, No. 1, Spring 2001
pp. 152–166

Innovation variety and patent breadth

Hugo A. Hopenhayn*

and

Matthew F. Mitchell**

When innovations are heterogeneous, it may be advantageous to provide a variety of patents. By trading off patent breadth for length, it is possible that fees are not needed in the optimal policy. We present two examples. The first is a quality-ladder model, in which innovations benefit society directly as well as through their use as building blocks to future inventions, and the rate of arrival for the future innovation is unobserved. More fertile innovations get more breadth for a shorter time. Menus may also be useful in the case of horizontal product differentiation.

1. Introduction

■ Patent policy is the centerpiece of many nations' attempts to encourage innovation by granting a property right to an inventor. In the language of modern economic theory, an inventor is given the right to exclude others from producing over a part of the product space. One element of the protection is the length of time for which the protection lasts. Another is the set of products that at any given time may be prevented by the patentholder: in other words, the patent's breadth. Patent protection is costly because it generates market power for the innovator; it is necessary because inventions are costly to produce but may be nonrival (costless to reproduce) after their invention, leaving the inventor without a means of benefiting without some protection.

This article is concerned with optimal patent design in the presence of innovation heterogeneity. A patent is defined by its breadth, its length, and its origination and/or renewal fee. These are indeed the instruments that have been considered in the literature and in the public policy debate. The problem is modelled here as an optimal mechanism design where innovators have private information about the nature of their innovations. This is the framework followed in recent articles; for instance, Cornelli and Schankerman (1999) and Scotchmer (1999) use asymmetric information to justify the use of patents as an optimal incentive scheme. We first consider an abstract reduced-form model that extends Gilbert and Shapiro (1990) to allow for heterogeneity. We introduce a sorting condition and show that the optimal patent menu has zero fees, trading off patent length for breadth. To illustrate the relevance of this abstract model, we consider two specific applications: a model of a quality ladder and one of spatial competition.

The quality-ladder model captures the idea that innovations are not only useful in themselves (or together with existing knowledge) but are also building blocks to future

* Universidad Torcuato Di Tella and University of Rochester; huho@troi.cc.rochester.edu.
** University of Minnesota; mitchell@atlas.socsci.umn.edu.
Thanks to Fernando Alvarez, April Franco, Nancy Gallini, Gerard Llobet, Robert Merges, Juan Pablo Nicolini, Andy Skrzypacz, two anonymous referees, and Editor Jennifer Reinganum as well as participants at the NBER Patents Conference, January 8–9, 1999. We are at fault for mistakes.

innovations. Heterogeneity arises as some may provide more ''fertile'' subsequent research than others. This quality-ladder model is similar to the one in O'Donoghue, Scotchmer, and Thisse (1998). A first innovator has an invention of given value and unobserved degree of ''spillovers'' to future projects. The spillovers take the form of the speed with which improvements to the product can be made. Higher arrival of improvements is a benefit to society, but it is a deterrent to patentees who find that their invention is more rapidly made obsolete. The fact that patents encourage investment through sales of the patented item means that fertile research areas, which lead to rapid improvements, are least profitable for the original inventor.[1] The patent authority, then, should provide specialized incentives to the patentee in fertile research areas. This takes the form of broader patents. But broad patents may be inappropriate in slower-growth areas; the patent office can, however, offer several varieties of patent protection in such a way as to induce innovators to choose the protection best suited to society. Just as in such classic works as Nordhaus (1969), the government seeks, on the one hand, to minimize the length of a patent because of the social cost of the monopoly right it confers, while also realizing that such a cost is necessary to reward innovators and encourage research.

The second application extends Klemperer's (1990) model of horizontal differentiation, introducing heterogeneity of imitation costs. In the absence of heterogeneity, Klemperer finds that the optimal patent should have infinite length, while breadth should be as narrow as possible. When heterogeneity is introduced, we find that even if a single patent is offered, where breadth is constrained to take a single value, the optimal patent will not involve infinite length. Moreover, as an implication of our abstract model, it can be efficient in this setup to offer a variety of patent breadths.

Our article is closely related to two separate areas in the patent design literature. On the one hand are the articles that consider explicitly the tradeoff between patent breadth and length. This literature either abstracts from heterogeneity or implicitly assumes that the patent office can perfectly discriminate among different innovation types. An important article in this area is Gilbert and Shapiro (1990), which studies a reduced-form design problem where the patent authority seeks to provide an inventor with a fixed amount of monopoly profits to exactly offset the costs of research. The articles by Klemperer (1990) and O'Donoghue, Scotchmer, and Thisse (1998) consider a similar design problem in the context of more specific structural models. Our article suggests that some of the results obtained in this literature may not be robust to the inclusion of private information.

On the other hand are the articles by Cornelli and Schankerman (1999) and Scotchmer (1999), which explicitly consider optimal patent design in the face of incomplete information on the part of the patent office. In these articles, innovations are sorted out by trading off the length of the patent against a fee.[2] In fact, since both U.S. and European patents are now based on a renewal-fee system, such a menu of lengths and fees is already in place, albeit at seemingly low fees. These articles restrict focus on fees and length as the instruments for patent design.[3] Our article shows that breadth is a valuable instrument that can indeed dominate the use of patent fees as a sorting device.

The article is organized as follows. Section 2 considers the abstract model and derives the main results on the role of length and breadth as sorting instruments. Section 3 considers the quality-ladder model of innovation fertility. Section 4 considers the

[1] This tension was first outlined in Green and Scotchmer (1995). Moreover, Scotchmer (1996) suggests that second-generation products should not be patentable at all.

[2] As noted by Scotchmer (1999), asymmetric information is necessary for patents to be an optimal incentive scheme to reward innovation.

[3] Also related is an article by de Laat (1997), which compares patents of a fixed breadth to a cash prize under the assumption that the patent office is not informed about the cost of an innovation.

horizontal-differentiation model. Finally, Section 5 provides some discussion about the implementation of patent menus.

2. The general design problem

■ Consider a patent authority faced with the following problem. An idea $\theta \in \Theta \equiv \{1, \ldots, J\}$ arrives to an innovator at time zero with probability $g(\theta)$. Translating ideas into inventions requires the inventor to combine his idea with a research cost c. Innovators may always choose to pay a cost of zero and create an ‘‘innovation’’ of zero value that cannot be distinguished from a valuable innovation by the patent authority. The patent authority also does not observe θ.

Innovators can make profits if and only if they are given a property right. This property right has breadth $B \in \mathbb{R}$ per unit of time and lasts for T periods. The innovator's profits are $\Pi(B, T, \theta)$. The function Π is continuous as well as increasing in the first two arguments. Since we assume property rights (B and T) are essential for the innovator to make profits, $\Pi(0, T, \theta) = \Pi(B, 0, \theta) = 0$. The authority may charge a fee F for this property right. Notice that since zero-value innovations are freely available and indistinguishable from the rest, the government will never be able to offer direct monetary compensation to innovations. This moral hazard problem restricts fees to be nonnegative.[4]

Society benefits from patents according to the continuous function $S(B, T, \theta)$, which is decreasing in the first two arguments to reflect the fact that market power is costly to society. Notice that we do not include F in the social welfare function. The assumption is that transfers from the producer to the government are welfare neutral. In a sense, this is equivalent to assuming that the government has access to a nondistortive tax instrument, a lump-sum tax, and therefore there is no social gain in transfers to the government. The reason for this assumption is so that it is never efficient for the government to ‘‘sell’’ patent power to raise revenue. It is not hard to imagine better revenue-raising instruments than patents; for instance, a consumption tax drives a wedge between price and marginal cost like monopoly power, but it does not affect future innovators. Assuming that there is a better revenue-raising instrument than this sale of monopoly power is enough to maintain the results.

The patent authority solves the problem

$$\max_{B(\theta),T(\theta),F(\theta)} \sum_{\Theta} S[B(\theta),\, T(\theta),\, \theta]g(\theta)$$

such that

$$\Pi[B(\theta),\, T(\theta),\, \theta] - c - F(\theta) \geq 0 \tag{IR}$$

$$\Pi[B(\theta),\, T(\theta),\, \theta] - c - F(\theta) \geq \Pi(B(\hat{\theta}),\, T(\hat{\theta}),\, \theta) - c - F(\hat{\theta}), \quad \forall \hat{\theta} \tag{IC}$$

$$F(\theta) \geq 0, \qquad \forall \theta. \tag{Moral Hazard}$$

The problem is formulated as an optimal revelation mechanism. The patent authority sets a menu of patents, with a breadth, length, and fee for each type θ. We will assume that all useful innovations are worthy of being implemented,[5] the first constraint (individual rationality) says that the menu grants to each type sufficient profits to cover costs. Since research costs are positive, some property rights will be conveyed. The

[4] This is a minimal setup for patents to arise as an optimal mechanism.

[5] This assumption is why it is justified to incorporate θ into the function S: all projects are implemented, and therefore society can value innovations according to the θ they will inevitably bring.

second constraint (incentive compatibility) ensures that each type θ accurately ''reports'' its type by choosing the appropriate menu item. The final constraint, moral hazard, ensures that worthless innovations are not patented.

This formulation is similar to Gilbert and Shapiro (1990), who study a complete-information version of this form. Because of the complete information, there is no need for fees. They use the special case $\Pi(B, T) = \int_0^T e^{-\rho t}\pi(B)\,dt$, where π are instantaneous profits as a function of B and ρ is the discount rate. The interpretation is that patents confer a constant reward, increasing in B, until the patent expires. The form used here allows for the possibility that the patent may effectively end before T due to obsolescence of the product. For shorthand, denote $B(\theta_j)$ by B_j, $T(\theta_j)$ by T_j, and $F(\theta_j)$ by F_j.

The use of renewal systems (a menu of fees and lengths) as a sorting mechanism has been studied in recent work on patent design with incomplete information. When breadth is added as an instrument, the optimal patent menu may involve a variety of breadths. In fact, using fees may be inferior to breadth as a sorting mechanism. Since patents reimburse fixed costs, and since the patentee views the fee as such a cost, charging a fee increases the costs that must be reimbursed with socially costly market power.

We now introduce a set of sorting conditions that will be sufficient for the optimal level of fees to be zero:[6]

Sorting conditions:

$\Pi_1(B, T, \theta)$ is strictly increasing in θ
$\Pi_2(B, T, \theta)$ is strictly decreasing in θ
$\Pi(B, T, \theta)$ is monotonic (increasing or decreasing) in θ.

The type space θ is taken to index projects by their profitability, either up or down. High-θ types value breadth more and time less. This sorting condition allows us to prove the following:

Proposition 1. Under the sorting conditions, there is an optimal patent menu (B_j, T_j, F_j) where $F_j = 0$ for all j.

Sorting with breadth is better than with a fee because it does not inflate the fixed cost to be reimbursed. But what is the interpretation of the sorting condition? In the following section we introduce a model consistent with the conditions. The flavor of the example is that the parameter θ indexes the degree to which a product leads to subsequent products. Patentees with high θ value breadth because they are especially concerned about the possibility that future ideas might eclipse the breadth of protection they have been given. On the other hand, they value increments of time less because they are more likely to lose their market power before T due to noninfringing competitors. Their effective patent length is likely to be limited by the early arrival of competitors outside of their patent's breadth.

It is interesting to note that the sorting condition and the restriction to nonnegative fees are responsible for our corner solution. Relaxing one assumption or the other could upset this result. If the patent office can monitor part of the R&D costs, it will be optimal to compensate the innovator for this investment: fees will be negative. However, provided our sorting condition holds, fees will still not be used for sorting purposes. Our result depends on the sorting condition as well as the assumption that breadth and length are essential. As an example, one may interpret the model of Scotchmer (1999) or of Cornelli and Shankerman (1999) as one where breadth has no effect on profits. Such a model satisfies a weak version of the sorting condition, but it violates either the continuity of profits as a function of breadth or the assumption that breadth is essential.

[6] Subscripts denote derivatives.

To see why sometimes it may *not* be possible to sort without fees, consider the framework of Cornelli and Shankerman (1999). In that case, an innovator has a research productivity that is unknown to the patent office. Highly productive researchers can make ''good'' innovations that generate high profits per period, and thus they particularly value long patents. However, highly profitable products might be more apt to attract imitators (as, for instance, in Gallini (1992)), and therefore they might benefit more from breadth. If high-productivity inventors value both length and breadth more than a low-productivity inventor, it may not be possible to induce truthful reporting with any kind of length-breadth tradeoff, so fees may be useful as a sorting instrument. Even if there can be some sorting using length-breadth tradeoffs, the menu may require fees.

In our proof we show that any contract with positive fees can be improved upon by lowering them. As fees are reduced, the sorting property of the patent menu is preserved by lowering either breadth, length, or some combination of both. If the sorting condition is not satisfied, it could be necessary to increase breadth or length for some types; for some social surplus functions this could result in lower welfare. The sorting condition implies that the types can be ordered by their marginal valuation of breadth, and that those who value breadth the most value time the least. This may seem like an unusual condition, which is why we explore examples in the next two sections of cases where it can be met. On the other hand, it is easy to imagine a particularly important product for which the value of patent length may be high (since it is such an outstanding achievement), and also that this market is particularly prone to attracting competitors, since it has such great profit potential, and therefore breadth is also especially valuable for this type. When a type values both instruments more than another type, it is not necessarily the case that one can sort with the two instruments, and the optimal menu might involve fees.

Our argument also assumes that Π is monotonic in θ. This assumption is important to the result because it implies that only one individual-rationality constraint will bind. Suppose Π is decreasing in θ, as in the vertical-differentiation model described in the following section. If $c(\theta)$ is monotonically decreasing, the sorting condition may not be enough to let B and T efficiently screen innovators, since it may be that the IR constraint binds more than once. To reimburse a high-cost invention, a large amount of protection, in terms of B and T, is required, since these types also make low profits. Without a fee on the patent designed for the high-cost inventions, it might be impossible to keep other inventors from claiming to have the high cost in order to receive the large protection. Fees might be useful to prevent this misreporting.

The interpretation of the assumption of a decreasing c would be that it costs the innovator more to make the innovation difficult to improve upon, a sort of built-in protection for the patentee. On the other hand, if $c(\theta)$ is increasing in θ, so that it is more costly to generate innovations that lead more readily to future projects, the proof will hold without change, as θ is exogenously drawn and the IR constraint would bind only for the highest θ.

Under the sorting conditions, the optimal contract has a specific form. Projects with higher θ get more breadth and less length, since that is the instrument they favor most. The rest of the properties of the optimal patent menu are straightforward.

Proposition 2. Under the sorting conditions, the optimal patent menu has the property that $\theta_j > \theta_k$ implies $B_j \geq B_k$ and $T_j \leq T_k$. Moreover, if Π is strictly decreasing (increasing) in θ, then the IR constraint binds only for the highest (lowest) θ.

In terms of a model where θ indexes the degree to which one innovation facilitates future inventions, this means that an innovator bringing about a particularly fertile area of innovation should receive (weakly) more breadth. Only one type can be reimbursed exactly, c. All others receive a bit more due to an informational rent, as in Scotchmer (1999).

Currently, there is a sort of patent menu: most patent systems, including the one in the United States, are renewal systems. The fees charged are a function of the length of protection granted. There is no menu of breadths in the statute, however. In the notation here, one can describe the protection provided as being a menu of patents $(B, T(\theta), F(\theta))$. Fees are positive for all θ. When the sorting condition holds, Proposition 1 implies that this is not optimal; however, that condition may not hold. But it is notable that a simple result in the finite-types model is that it is never optimal to have $F(\theta) > 0$ for all θ, regardless of the sorting condition. The reason for this is clear: the patent office could do better by lowering everyone's fees, thus eliminating any problems with the IR constraint, and then lower a costly instrument for the type which favors that instrument, at the margin, the least.

3. Breadth and innovation fertility: a sequential model

■ The model above takes as given a reduced-form profit function of the potential patentee, in the spirit of work by Gilbert and Shapiro. It is useful to consider a more structural model, where the profit function is derived from demand and a specification of the technology. We consider the case of an innovation that arrives and then may be superseded, so that the patent's effective length (the time until a noninfringing improvement drives the original innovator out of the market) is different from its statutory life. The model is similar to that in O'Donoghue, Scotchmer, and Thisse (1998); it is simplified to allow for only two innovations (they consider many), but the arrival rate is unknown.

□ **Environment.** Consider a quality ladder and a continuous, infinite-time horizon discounted by ρ. Before the first innovation arrives, a technology producing a good of quality normalized to zero is freely available. The marginal cost of production of all qualities will be taken to be zero, for simplicity. The inventor arrives with an observable improvement π, which also has an unobservable parameter θ, which we take to be uncorrelated with π. A second innovation arrives according to a Poisson process with arrival parameter θ; that is, the higher is θ, the sooner is the expected arrival of the second innovation. This is the sense in which high θ means that the first patentee has developed a technology that is fertile: it leads to improvements sooner.

When the second innovation arrives, its quality *improvement* over π is given by Δ, which is distributed according to a continuously differentiable cumulative density $H(\Delta)$ with support on $\mathbb{R}_+$. For simplicity, we assume the second innovation can be implemented at no cost, so that its inventor wishes to implement it regardless of the patent protection it might receive. The patent authority is charged with designing a patent menu for the first innovator. The menu is, as above, $B(\theta)$, $T(\theta)$, $F(\theta)$, where $B(\theta)$ is the breadth of the patent. The interpretation is that when the second innovation arrives, it can be produced provided that it has $\Delta > B(\theta)$. T is the length of time the protection is conferred; after that, any improvement can be produced, and the patentee's product can be freely produced by competitive firms, so that profits fall to zero after T. For simplicity, there are only these two ideas, and not an infinite series of improvements. Denote by $P(t, \theta)$ the probability of an arrival of a second innovation in the first t periods for an innovation of type θ. Since the distribution H is exogenous, this is a model about which ideas are implemented, and not one about where the ideas ‘‘come from.’’[7] Moreover, our analysis is limited by the fact that we optimize only

[7] Clearly it would be feasible to make when and how big the improvement is something the second firm could influence. The way that research affects the arrival of ideas is a question that has been posed in, for instance, Horowitz and Lai (1996).

over implementing the current project. We do not consider the distortions created by applying this policy to many innovations, arriving in sequence.[8]

Next, we describe the static competition game. Both the patentee and the second innovator simultaneously choose a price. The quality 0, which is freely available, is sold at marginal cost, $p = 0$. There is a single representative consumer who demands a single unit and has reduced preference

$$u(q, p) = q - p,$$

where q is the quality of the good purchased and p is the price paid. When the second innovation is not involved, the equilibrium has $p_\pi = \pi$ and the patentee earns π units of profit. When the second innovation has arrived and is sufficiently different to be allowed to produce, the equilibrium has $p_\pi = 0$ and $p_\Delta = \Delta$. Consumers are indifferent between the two products; it is assumed that they buy from the highest quality and so the patentee earns profits 0 and the second innovator earns profits Δ.

The Coase theorem suggests that it is efficient to grant large patent power if efficient licensing agreements can be made. In this context, note that if the second innovator can buy the first patent at its value to the patentee, he can always be assured of all the incremental profit of his innovation by paying the patentee all of the future monopoly profit flows. Since the monopoly pricing in this problem leads the monopoly provider to extract all the consumer surplus from the agent, the second innovator's decisions are undistorted by large monopoly power to the first agent.

Because we seek to investigate the social costs of patent power as including distortions to future innovators, we do not allow for such patent buy-out. In particular, we assume that innovations $\Delta < B$ are not implemented until the original patent expires due to overline power of the patentee. This means that the cost of patent power has the effect of slowing second-generation innovations. On the other hand, there is no cost to the monopoly power that a patent provides: given the setup, the patentee extracts all the surplus from the consumer. Later we investigate cases where this does not hold, and we confirm that it is not crucial to the results.

□ **Optimal patent design.** We can now write this model in the form of the previous section by characterizing Π and S. It turns out that the economy satisfies the sorting conditions, and thus $F = 0$ in the optimal patent menu. This is interesting for two reasons: First, it is plausible that the optimal mechanism in reality should involve no fees (or perhaps less prevalent fees), but rather a tradeoff of length for breadth. Second, it makes the solution to the design problem simpler to characterize and more intuitive, since effectively it has only two instruments.

The patentee makes profits π until a second innovation arrives and the innovation is bigger than B, at which point the patent effectively expires. Discounted expected profits for the patentee are

$$\Pi(B, T, \theta) = \int_0^T e^{-\rho t}\pi[(1 - P(t, \theta)) + P(t, \theta)H(B)]\, dt.$$

Proposition 3. The sorting conditions hold for the sequential model.

[8] Llobet, Hopenhayn, and Mitchell (1999) explore optimal breadth and length with asymmetric information and many innovations but a constant fertility.

The statutory length of the patent T matters to the patentee only if no innovation arrives that makes the initial innovation obsolete. That is, T is an option for additional monopoly power if the patent has not been superseded. The higher is θ, though, the less valuable is that option, since it is more likely that an arrival will have occurred that has made the additional T moot. Breadth is relevant when and if another innovation arrives. The faster are arrivals (higher θ), the more valuable breadth is, since it is more likely to be useful and likely to be useful sooner. The sequential model satisfies the sorting conditions introduced above because fast arrival means that B is more valuable and T is less valuable.

While the model points out an environment where the sorting conditions hold, it also points to reasons why they may not. For instance, suppose that π and θ are positively correlated and π is unobserved. It may not be that higher θ prefer more breadth and less time of protection, since each instant tends to bring more profits for the high-θ firm. When the sorting conditions fail, it may be optimal to incorporate fees for some patents even though this increases the costs of the innovator.

The patent authority maximizes social gain, the sum of producer and consumer surplus. Before the second innovation arrives, the patentee earns π and consumers get zero utility. If an improvement has arrived but is less than B, the patentee retains monopoly and nothing changes. When an idea arrives and either it is greater than B or the patent has expired, it can be produced. It is sold at a price of Δ and gives the consumer surplus π, so social gain is $\pi + \Delta$. In sum we have

$$S(B, T, \theta) = \left\{\pi \int_0^\infty e^{-\rho t}\, dt + E(\Delta \mid \Delta > B)[1 - H(B)] \int_0^T e^{-\rho t} P(t, \theta)\, dt \right.$$
$$\left. + E(\Delta) \int_T^\infty e^{-\rho t} P(t, \theta)\, dt \right\}.$$

The cost of patents is that they lead to slower arrival of the second-generation innovation. Society always enjoys π, either from profits or consumer surplus. The second term includes the incremental social gain from an innovation that arrives by time T and is implemented. Until T the innovation is implemented only if its improvement Δ exceeds B, in which case it provides an expected social gain equal to the expectation of Δ conditional on $\Delta > B$. The last term includes the fact that after the patent expires at time T, the improvement is always implemented.

Next, by way of an example, we show circumstances under which the menu is nontrivial, i.e., it is not only incentive feasible to trade B for T in the optimal mechanism, it is also optimal given S.

□ **An example.** To illustrate the role and optimality of nontrivial patent menus, consider a simple case where $H(\Delta)$ is a point mass on some $\overline{\Delta}$. There are two types of original projects: θ_1, which is low (slow arrival), and θ_2, which is high (fertile). Suppose that

$$\int_0^{T_1} e^{-\rho t} \pi [1 - P(t, \theta_1)]\, dt = c$$

for some finite T_1. In this case, a T_1-period patent of breadth zero is sufficient to induce investment for potential patentees θ_1. The patent is simply conferring the monopoly right to sell the project through T_1. Improvements are free to infringe. Since the price-setting game implies that monopolists extract all the surplus from consumers, this sort of monopoly power has no social cost, that is, it maximizes $S(\,)$ without constraint. It is straightforward to extend the model to cases where monopoly rights have social cost and maintain the results here; one such model is discussed below.

On the other hand, suppose that θ_2 is very high, so that the return to an infinite-time, zero-breadth patent for θ_2 is not sufficient to offset c:

$$\int_0^\infty e^{-\rho t}\pi[1 - P(t, \theta_2)]\, dt < c.$$

Without breadth, fertile ideas θ_2 are not implemented, even if $T_2 = \infty$. For instance, if θ_2 nears infinity, discounted profits are near zero, since an improvement arrives almost immediately. Fertile innovations require breadth.

This example illustrates the tradeoff faced by the patent office. It is forced to provide breadth $B_2 = \overline{\Delta}$ to induce innovation in very valuable, high-fertility areas; however, this protection is very costly when $\overline{\Delta}$ is large, and so it is used only when absolutely necessary to induce the original project. The length of the protection T_2 is chosen to solve

$$\int_0^{T_2} e^{-\rho t}\pi\, dt = c.$$

Since $P(t, \theta_1) \in (0, 1)$, $T_2 < T_1$. The patent authority must provide (complete) protection for the fertile innovation but can provide it for a shorter interval. Notice that offering this patent $(\overline{\Delta}, T_2)$ to the fertile type and the patent $(0, T_1)$ to the low type is incentive compatible: both give the low type θ_1 the same reward, and patentees of fertile inventions strictly prefer $(\overline{\Delta}, T_2)$. The patent authority, then, can screen by offering these two types of patents, and in fact finds it optimal to do so. Fertile innovations get protection from future projects; infertile areas get monopoly rights for a longer time interval but no right to exclude significant improvements.

4. Horizontal differentiation, imitability, and patent breadth

■ This section considers a horizontal-differentiation model as in Klemperer (1990), where innovations vary in the extent to which they can be imitated. It is a Hotelling-style model, with a continuum of goods. There is a unit mass of consumers for the patented good. Each has a reservation price $\overline{p}$ and a cost of substitution τ per unit in the product space away from the patented good. Each consumer chooses to consume one unit or none. They may consume either the patented good, which has price p and zero marginal cost, or a competitively produced substitute B units away, which is priced at marginal cost m and so has effective price $m + \tau B$. Klemperer (1990) shows that under plausible conditions, it is optimal to be at a corner, i.e., maximum time of patent life or maximum patent breadth. When there is unobserved heterogeneity, though, not only might that not be the case, it might also be useful to screen innovators using multiple patent breadths.

In the sequential-innovations model, θ indexed the speed of new arrivals. Here θ will affect the cost of the competitor. Suppose that none of the products is possible before the arrival of the potentially patented invention. After the arrival, if the innovator chooses to invest c, the good that is invented can be produced at cost zero and the rest of the goods can be produced at a cost $m(\theta)$. Let $m(\theta)$ be a decreasing function, so that high θ indicates low-cost substitutes are available. Inventions with high θ are more flexible in the sense that they lower costs by a greater amount compared to other possible products. This is similar to the role of θ before: higher θ means more competition faces the patentee.

Consider the case (considered by Klemperer) where τ is homogeneous across consumers and $\bar{p}$ is heterogeneous, distributed according to $D(p)$. The cost of patent breadth, in this case, is that the patentee may price some consumers out of the market in order to extract surplus from consumers with more willingness to pay. Klemperer shows that with complete information by the patent office, it is optimal to provide a long-lived patent of as narrow a breadth as possible to reimburse c, if c is small enough. With incomplete information about some characteristics of the patented good, though, this may not be true.

An alternative way to view τ when it is homogeneous is as a per-unit cost of producing a competing good, similar to the concept of designing around. The idea is that a competitor must spend resources to modify the good sold so that it does not infringe. It is natural to think of breadth as increasing this cost for potential competitors.

In a given period, the patentee solves

$$\pi(B, \theta) = \max_{p \in [0, m(\theta) + \tau B]} pD(p).$$

The patentee must price below $m(\theta) + \tau B$; otherwise, no one buys the patented good and they all either substitute or do not consume. Given that constraint, no consumer buys a substitute good. Inefficiency arises because the monopolist's profit-maximizing price p^* may be positive and, hence, greater than marginal cost. Letting d denote the derivative of D, the social surplus is

$$S(B, T, \theta) = \left(\int_0^\infty pd(p)\, dp\right)\left(\int_0^\infty e^{-\rho t}\, dt\right) - \left(\int_0^{p^*} pd(p)\, dp\right)\left(\int_0^T e^{-\rho t}\, dt\right).$$

The second term reflects the cost of excluding, for the first T periods, agents with valuations below p^*.

An invention of a given type can earn profits

$$\Pi(B, T, \theta) = \pi(B, \theta) \int_0^T e^{-\rho t}\, dt.$$

To the patentee, competitors constrain the set of feasible prices. The following proposition establishes a sufficient condition for the optimal level of fees to be zero. Although only a weak version of the sorting conditions holds, the basic argument of Proposition 1 works in this special case.

Proposition 4. Suppose $pD(p)$ is strictly concave. Then the optimal patent menu for the horizontal-differentiation model has $F = 0$.

The higher is θ, the more important are competitors, and, in turn, the lower are profits per instant. Therefore, the value of having monopoly power for a longer time decreases with θ. The value of B is the shadow value of increasing the constraint in the firm's profit-maximization problem at the rate τ. If the constraint does not bind, B has no marginal value to the patentholder. When the constraint binds, the sign of Π_{13} depends on properties of d. If those properties are satisfied, the sorting conditions apply.

□ **What if only one patent is offered?** One point to emphasize is that even if one restricts the patent authority to offer a single type of patent, that patent is not necessarily infinitely lived, in contrast with Klemperer's result. The reason is that here, unlike with complete information, the amount of profits is endogenous (except for the highest type, which will be on the IR constraint). By offering a broader but more short-lived patent, society loses relative to a long-lived patent that offers the same profits, but it gains to the extent that it allows the patent authority to deliver less monopoly profits to some types. This is because not every individual is being reimbursed for the costs of research; many are rebated more, due to rents from their private information.

Proposition 5. Consider the horizontal-differentiation model, and let c be sufficiently small and only one type of patent (B, T) be offered. It is not necessarily true that the optimal patent has $T = \infty$.

More generally than simply in this model, unobserved heterogeneity affects the optimal patent.

□ **Patent menus: an example.** Consider a case with two types, $\theta_L < \theta_H$, where c is small, and where the sorting conditions hold. The question is whether the patent office wants to offer two types of patents or just one. Suppose it is the case (proved to exist by the previous proposition) that an infinitely lived patent is not optimal if only one patent is offered. We will show that a menu of patents dominates a fixed patent $(\overline{B}, \overline{T})$ offered to all types. Consider offering θ_L a patent with infinite length that provides the same profits as alternatives on the menu, i.e., $(B(\theta_L), T(\theta_L)) = (\tilde{B}, \infty)$, where $\tilde{B}$ solves

$$\int_0^\infty \pi(\tilde{B}, \theta_L)e^{-\rho t}\, dt = \int_0^{\overline{T}} \pi(\overline{B}, \theta_L)e^{-\rho t}\, dt.$$

Under the sorting condition this is incentive compatible, since replacing B with T is most desirable for θ_L. For θ_H, then, nothing changes. Focus on θ_L. By Klemperer's proposition 2, conditional on a given reward, it is maximal to offer T as high as possible. That is, $S(\tilde{B}, \infty, \theta_L) > S(\overline{B}, \overline{T}, \theta_L)$, and the menu improves social welfare when $\theta = \theta_L$ without changing θ_H, an improvement over offering a constant $(\overline{B}, \overline{T})$ to all innovations.

This section, then, establishes two points: First, even without a menu of patents, the existence of heterogeneity affects optimal patent policy relative to cases considered in other work on optimal patent breadth. Second, patent menus may be optimal for sorting among projects with differing degrees of "flexibility" to other horizontally differentiated products.

5. Discussion

■ Patents reward innovators, but at a cost to society. In addition to the common monopoly pricing inefficiencies, patent breadth may retard the innovative activity it was meant to promote through the power it provides to initial innovators. To the extent to which different innovations provide different contributions to future research, a "one size fits all" patent policy is inappropriate. It may not provide sufficient protection to very valuable inventions that lead readily to second-generation products. They may provide the wrong sort of protection to various innovations. This intuition is true in a wide variety of cases, including both vertical and horizontal competition.

In some cases, the optimal response to this heterogeneity is to provide a menu of patent alternatives. This can be accomplished even when the patent authority has no information about the characteristics of patent applicants, and therefore it must rely on a revelation of types by the patentees. To the extent that breadth can be used as an instrument for protection and sorting, it will be preferable to sorting with a patent fee.

Even when a single patent is offered, its design should take into account the underlying heterogeneity. As pointed out in the article, it may be misleading to design a patent for one type of innovation in isolation without considering the multitude of different innovations that will be protected by that same patent policy. A patent that is optimal for a high-profit invention may not be sufficient encouragement for low-profit inventions; on the other hand, encouraging low-profit innovations can be detrimental when that same patent is owned by a higher-profit innovation.

An interesting point for future study is that when the patent statute is limited to a single definition of breadth, innovators may adopt other strategies in an attempt to gain more breadth. A commonly discussed tactic is the building of a ''patent wall,'' where the innovator simultaneously patents not only the base innovation but also a variety of related innovations, not so much for the value of those patents unto themselves, but rather to protect the central innovation. Such tactics can obviously be wasteful from society's standpoint; understanding more about how firms respond in the face of current policy might be a way to understand how important offering multiple patent breadths might be.

In our model we assume that all relevant information about the innovation is private to the original innovator and can be revealed only through a choice of menu. In practice, direct evidence on the nature of the improvement may be gleaned as the patent lives on.[9] As a consequence, courts may treat different innovations differently, in terms of breadth: fundamental breakthroughs are often given more breadth than marginal improvements. Of course, if courts could eventually ascertain everything relevant about a project, then it is trivial to ''sort,'' since punishments can be assigned so that truth telling is essential. The question of what can be gained by conditioning patent rights on the experience of the patented product is one that seems interesting, but it requires serious thought to be put into the question of what is known to whom, and when, since even eventual complete information can make the sorting problem trivial. Here we have taken the extreme assumption that θ is unknown to the patent office forever.

An important simplifying assumption used in the quality-ladder example is that there is only one improvement, not a string of improvements. One could imagine casting a problem such as this one in the full infinite-horizon setting of, for instance, O'Donoghue, Scotchmer, and Thisse (1998), using their definition of equilibrium and so on. The patent office, then, would have to set policy based on a sequence of reports up to time T. Since θ might be correlated from invention to invention, the patent authority might seek to use the reports in combination to extract the truth at less cost. Setting up a dynamic patent menu that takes advantage of the combined knowledge of the patent office from the reports it has received is beyond the simple subject matter described here; nonetheless, it seems that regardless of the structure, it will always be true that it may be useful to separate types with breadth menus, and that in a world with obsolescence, those who favor breadth (inventors who perceive obsolescence as nearby) may, for that same reason, not value statutory length, and therefore breadth-length menus can be an effective tool to sort innovators.

Patent breadth is an inherently vague concept in practice. The policy proposal of offering multiple patent breadths may seem, in light of this vagueness, particularly

[9] Such information is often revealed in the process of litigation. For an interesting article that introduces litigation as part of the patent design problem, see Llobet (1999).

difficult. However, there are a variety of ways to use the existing case law to redefine breadth.[10] One clear example is the doctrine of equivalents. In the famous *Graver Tank* case, the Supreme Court held that explicit infringement of a claim was not required to find infringement, that it was sufficient for the infringing innovation to ''work in substantially the same way, and accomplish the same result.'' This doctrine of equivalents has a range of applications. One recent case in which the Supreme Court upheld it, *Warner Jenkinson Company, Inc.* v. *Hilton Davis Chemical Co.,* illustrates its role. The case involved procedures for ultrafiltration of dyes. Before an innovation by Hilton Davis, the filtration required pH levels higher than nine. Hilton Davis introduced a method that, according to the claim of their patent, would operate at pH levels between six and nine.[11] Warner Jenkinson filed for a patent a year later, claiming a technique suitable for a pH of five. Hilton Davis claimed that its patent should extend as low as five by the doctrine of equivalents, despite the fact that (pH is a logarithmic scale) its claimed process was ten times as basic as the Hilton Davis procedure. In its ruling, the Supreme Court specifically stated that ''Congress can legislate the doctrine of equivalents out of existence at any time it chooses.'' Although the doctrine may not have been well defined at the time of the writing of the original patent statute, it is now an element of the law that can be defined. As a result, it would be possible to write a patent law with a menu of choices including allowing the patentee to decide whether or not to commit to having doctrine of equivalents protection, perhaps at the cost of a shorter patent length. The doctrine of equivalents is only one of many ways to describe breadth more precisely and to allow it to vary across patents.

There may be alternative equivalent ways of implementing patent breadth/length menus that do not require explicit consideration of breadth. In practice, many patents are licensed to subsequent innovators. Licensing or buyout fees are not typically a matter of regulation, yet they affect the set of second-generation improvements that actually take place. A limit to buyout prices can thus effectively implement breadth. This suggests that one way to offer varying degrees of breadth is to have the menu offer different limits on buyout fees. (See Scotchmer (1996) and Llobet, Hopenhayn, and Mitchell (1999)). Given that it may be advantageous to offer a menu of patent breadths, these alternative ways of implementing patent breadth seem to be a fertile area for future research.

Appendix

■ Proofs of Propositions 1–5 follow.

Proof of Proposition 1. First we state the following:

Definition. The (j, k) constraint is the IC constraint for the type θ_j agent with respect to reporting type θ_k, i.e., the constraint $\Pi(B_j, T_j, \theta_j) - c - F_j \geq \Pi(B_k, T_k, \theta_j) - c - F_k$.

Lemma A1. Let $\{b_j, t_j, F_j\}_{j=1}^{J}$ be an optimal patent system. If $F_j > 0$, there must exist $k < j$ and $\ell > j$ such that the k, j and ℓ, j constraints bind. Furthermore, $F_1 = F_J = 0$.

Proof. Suppose that $F_j > 0$ and that no constraint k, j for $k < j$ binds. Then consider the following alternative patent for type j: reduce b_j and F_j in such a way that the utility of type j remains unchanged. For all types greater than j, this represents a decrease in the utility associated to the j-patent. Furthermore, provided the changes are small, all incentive constraints for types lower than j will be satisfied. Since breadth is decreased for type j without changing patent length, the patent system thus obtained is superior, contradicting the optimality of the original one. The same procedure can be applied, instead reducing patent length, if no constraint k, j for $k > j$ binds. Finally, note that the argument applied implies immediately that for the extreme types 1 and J, patent fees must be zero. *Q.E.D.*

---

[10] For a more complete description of the patent law doctrine, see Merges and Nelson (1994).

[11] Interestingly, Hilton Davis originally filed for a patent where the claim simply said that its process worked for pH values less than nine. The patent examiner required the claim to be made more specific.

Lemma A2. Let $\{b_j, t_j, F_j\}_{j=1}^{J}$ be an optimal patent system. Let $J > j > 1$ and suppose $F_j > 0$. If for type j the constraint (j, k) binds, then $(b_k, t_k) \geq (b_j, t_j)$ and $F_k > 0$.

Proof. Assume (j, k) binds. There are three possibilities: either $(b_k, t_k) \geq (b_j, t_j)$, $(b_j, t_j) \gg (b_k, t_k)$, or the two vectors are not ordered. If the two vectors are not ordered, then if $j > k$ $(j < k)$, all types $\ell < j$ $(\ell > j)$ must strictly prefer the k-patent (ℓ-patent) to the j-patent, so by Lemma A1 F_j cannot be positive. Second, notice that $(b_j, t_j) \gg (b_k, t_k)$ would contradict optimality, for in such case the j type should be offered the k-patent. Consequently, $(b_k, t_k) \geq (b_j, t_j)$, and it must then be the case that $F_k \geq F_j > 0$, which completes the proof. *Q.E.D.*

Main Proof of Proposition 1. For any $j = 1, \ldots, J$, let $P(j)$ be the set of patent contracts to which j binds. Let A be the set of types that are offered patents with strictly positive fees. Suppose $A \neq \phi$. By Lemma A2, $P(A) \subset A$. But this implies that there is an incentive-compatible patent menu where all fees in A can be reduced. Such a patent menu could never be worse than the given one and by an appropriate reduction in breadth or length is potentially better. *Q.E.D.*

Proof of Proposition 2. From the sorting condition, since $F = 0$ for all θ, any time that $\theta_j > \theta_k$ and either $B_j < B_k$ or $T_j > B_k$, the IC constraints must be violated for either j or k. That the IR constraint binds for only one type is trivial from the monotonicity of Π in θ: if J is the lowest-profit type and another type is j, $\Pi(B_j, T_j, \theta_j) - c > \Pi(B_J, T_J, \theta_J) - c \geq 0$. *Q.E.D.*

Proof of Proposition 3. The patentee makes profits π until a second innovation arrives and the innovation is bigger than B, at which point the patent effectively expires. Discounted expected profits for the patentee are

$$\Pi(B, T, \theta) = \int_0^T e^{-\rho t}\pi[(1 - P(t, \theta)) + P(t, \theta)H(B)]\, dt. \tag{A1}$$

Although Θ is the set of integers 1 to J, it is sufficient for the sorting condition that the cross derivatives of Π, taking θ to lie in the real line, have the appropriate signs. Notice that

$$\Pi_{23}(B, T, \theta) = e^{-\rho T}\pi[-(1 - H(B))P_2(T, \theta)] < 0,$$

since $P_2 > 0$.

The derivative of the integrand in (A1) with respect to B is $e^{-\rho t}\pi[P(t, \theta)h(B)]$, where h is the density function of H. Taking the derivative of that with respect to θ yields $e^{-\rho t}\pi[P_2(t, \theta)h(B)] > 0$, which implies the cross derivative $d^2/(dBd\theta)$ of the integrand is positive. As a result, $\Pi_{13}(B, T, \theta) > 0$. *Q.E.D.*

Proof of Proposition 4. Define $\overline{B}_i$ to be the breadth that solves $m(\theta_i) + \tau\overline{B}_i = p^M$, where p^M is the monopoly price $p^M = \text{argmax } pD(p)$. Define $\pi^M = p^M D(p^M)$. $\overline{B}_i$ is increasing in i, since $m(\theta)$ is decreasing. Define $p_i(B) = m(\theta_i) + \tau B$ to be the price when the constraint in the patentee's problem binds. Note that p_i is strictly decreasing in i.

Whenever the constraint on the patentee's problem binds, the sorting conditions are immediate. The marginal value of time is $p_i(B)D(p_i(B))$, which is strictly decreasing in i. The value of breadth is $d\{p_i(B)D[p_i(B)]\}/dp$, which is strictly increasing in i by strict concavity.

Type i does not value additional breadth for $B \geq \overline{B}_i$, and the marginal value of T is π^M whenever $B \geq \overline{B}_i$. In those cases, the monotonicity of the derivatives of Π is true only weakly. We can restrict attention to $B_i \in [0, B_i]$, since it is irrelevant (and potentially costly in terms of incentives for the other types) to offer any more.

Consider the proof of Proposition 2. Lemma A1 is unchanged. The following lemma replaces Lemma A2.

Lemma A3. Let $\{b_j, t_j, F_j\}_{j=1}^{J}$ be an optimal patent system. Let $J > j > 1$ and suppose $F_j > 0$. If for type j the constraint (j, k) binds, then $F_k > 0$.

Proof. Nothing in the argument for Lemma A2 changes except for the following:

"If the two vectors are not ordered, then if $j > k$ $(j < k)$, all types $\ell < j$ $(\ell > j)$ must strictly prefer the k-patent (ℓ-patent) to the j-patent, so by Lemma A1 F_j cannot be positive."

Suppose $j < k$. Then it is impossible that $\overline{B}_\ell \leq B_j \in [0, \overline{B}_j]$, and so type ℓ strictly prefers the k patent, and the argument of Lemma A2 goes through.

On the other hand, consider the case where $j > k$. The possibility that is incentive compatible is $B_j > B_k$ and $T_j < T_k$. Now it is possible that for $\ell < j$, type ℓ only weakly prefers k to j. Suppose some type ℓ is indifferent between k and j. This can be the case only if $B_j > B_k \geq \overline{B}_\ell$, in which case type ℓ gets monopoly profits for T periods under each patent. Therefore

$$F_j - F_k = \pi^M \int_0^{T_j} e^{-\rho t}\, dt - \pi^M \int_0^{T_k} e^{-\rho t}\, dt.$$

Since $T_j < T_k$, $F_j - F_k < 0$, so $F_k > F_j \geq 0$. *Q.E.D.*

That fees are zero is now a consequence of an identical argument to Proposition 1. *Q.E.D.*

Proof of Proposition 5. Consider the case where there are two types, θ_L and θ_H. If the government is offering only one patent type and wants to have an infinitely lived patent, the breadth solves

$$c = \pi(B, \theta_H) \int_0^{\infty} e^{-\rho t}\, dt$$

if it is to encourage θ_H. For sufficiently high $m(\theta_L)$, this B implies that the constraint $p \leq m(\theta_L) + \tau B$ does not bind, i.e., the type-θ_L patentee enjoys monopoly power forever under the proposed patent. If c is small enough, the costs associated with any patent (B, T) that reimburses exactly c is small, yet the costs of monopoly power for θ_H forever are not. Reducing the length of time for which θ_L's monopoly lasts, together with increasing B to satisfy the IR constraint of θ_H, is always welfare improving. *Q.E.D.*

References

CORNELLI, F. AND SCHANKERMAN, M. ''Patent Renewals and R&D Incentives.'' *RAND Journal of Economics,* Vol. 30 (1999), pp. 197–213.

DE LAAT, E.A.A. ''Patents or Prizes: Monopolistic R&D and Asymmetric Information.'' *International Journal of Industrial Organization,* Vol. 15 (1997), pp. 369–390.

GALLINI, N. ''Patent Policy and Costly Imitation.'' *RAND Journal of Economics,* Vol. 23 (1992), pp. 52–63.

GILBERT, R. AND SHAPIRO, C. ''Optimal Patent Length and Breadth.'' *RAND Journal of Economics,* Vol. 21 (1990), pp. 106–112.

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

HOROWITZ, A.W. AND LAI, E.L.C. ''Patent Length and the Rate of Innovation.'' *International Economic Review,* Vol. 37 (1996), pp. 785–801.

KLEMPERER, P. ''How Broad Should the Scope of Patent Protection Be?'' *RAND Journal of Economics,* Vol. 21 (1990), pp. 113–130.

LLOBET, G., HOPENHAYN, H., AND MITCHELL, M. ''Rewarding Sequential Innovations: Patents, Prizes, and Buyouts.'' Mimeo, University of Rochester, 1999.

MERGES, R.P. AND NELSON, R.R. ''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.

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

O'DONOGHUE, T., SCOTCHMER, S., AND THISSE, J-F. ''Patent Breadth, Patent Life and the Pace of Technological Progress.'' *Journal of Economics and Management Strategy,* Vol. 7 (1998), pp. 1–32.

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

———. ''On the Optimality of the Patent System.'' *RAND Journal of Economics,* Vol. 30 (1999), pp. 181–196.