NBER WORKING PAPER SERIES

INNOVATION FERTILITY
AND PATENT DESIGN

Hugo A. Hopenhayn
Matthew F. Mitchell

Working Paper 7070
http://www.nber.org/papers/w7070

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 1999

Presented at the NBER Conference "The Patent System and Innovation" funded by the Alfred P. Sloan Foundation via a grant to the NBER Project on Industrial Technology and Productivity. Thanks to Fernando Alvarez, April Franco, Nancy Gallini, Gerard Llobet, Robert Merges, Juan Pablo Nicolini, and Andy Skrzypacz, as well as to Adam Jaffe for helping get access to the patent database. We are at fault for mistakes. The views expressed in this paper are those of the authors and do not reflect those of the National Bureau of Economic Research.

Innovation Fertility and Patent Design
Hugo A. Hopenhayn and Matthew F. Mitchell
NBER Working Paper No. 7070
April 1999

ABSTRACT

It may be advantageous to provide a variety of kinds of patent protection to heterogenous innovations. Innovations which benefit society largely through their use as building blocks to future inventions may require a different scope of protection in order to be encouraged. We model the problem of designing an optimal patent menu (scope and length) when the fertility of an innovation in generating more innovations cannot be observed. The menu of patent scope can be implemented with mandated buyout fees. Evidence of heterogeneous fertility and patent obsolescence, keys to the model, are presented using patent data from the US.

Hugo A. Hopenhayn
University of Rochester
Harkness Hall
Rochester, NY 14627-0156

Matthew F. Mitchell
Mgmt. and Econ. Building
University of Minnesota
271 19th Avenue South
Minneapolis, MN 55455

1. Introduction

Patent policy is the centerpiece of many nations' attempts to encourage innovation. Since innovation is thought to be central to both macroeconomic performance and the microeconomic structure of industry, the study of patent policy is a central topic in policy discussions of the economics of innovation and growth. Patents are the granting of 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 which, at any given time, may be prevented by the patent holder. Patent protection is costly because it generates market power for the innovator; it is necessary because inventions are costly to produce but may be nonrivalrous (costless to reproduce) after their invention, leading the inventor without a means of benefiting without some protection.

Innovations can have many benefits. One is that they may be useful in themselves (or together with existing knowledge) in producing something directly. Also, innovations are building blocks to future innovations, and some may provide more "fertile" subsequent research than others. Both characteristics are valuable, but they might require quite different sorts of patent protection. Patents reward innovators through monopoly profits; very fertile innovations may have a hard time capitalizing on this right if it is defined narrowly so that subsequent projects erode the original innovator's market power too readily. Specifically, for a given level of protection, profits may be *decreasing* in the usefulness of the invention to future innovators. Broad protection, in the sense of protecting against a wide variety of subsequent innovations, may be required to encourage such innovations. On the other hand, protection against subsequent innovations clearly discourages future projects, and thus deters innovation. Finding the correct balance depends, then, may depend on details of the invention, details that may be hard for the patent authority to ascertain. We investigate the potential for patent policy to provide different protection for different patents, considering the fact that it may be impossible for the patent authority to observe the differences in fertility ex-ante when the patent is issued. The patent describes the set of products protected in a given period (Breadth) and the length of time of the protection, and the patentee chooses a breadth/length combination from a patent "menu" when the patent application is filed.

The literature on optimal patent design has focused on at least two areas that are directly relevant to this work. Several papers have studied the problem as one of providing some breadth of protection for some length of time. Gilbert and Shapiro (1990) study 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. A more structural model is the one in O'Dognohue et al. (1997). Innovations arrive along a quality ladder, and breadth of protection is defined in terms of what kinds of future projects are within the scope of the original patent. They find that, in a world where improvements arise more frequently, the socially optimal amount of patent breadth should be greater than if there is less potential for future projects. The question asked here is how patent policy can be made to do this when fertility of innovations is unobserved.

Other papers have tackled some problems of patent policy in the face of incomplete information on the part of the patent office. In those papers, the focus has been on sorting products by their usefulness by trading off the length of the patent against a fee. In fact, since both US 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. Cornelli and Schankerman (1997) and Scotchmer(1997) show, however, that in a world where patent characteristics are unobserved to the patent office, it can be optimal to sort different types of innovations into different types of patents. The different types of patents they focus on are patents of different lengths; the difference in this paper is that both the length and the breadth of the patent are used as instruments. This is consistent with work such as Klemperer (1990) and Gilbert and Shapiro (1990) who consider the use of both as instruments.

For the most part, we work with a sequential "quality ladder" model similar to the one in O'Dognohue et al. (1997). Our model has several key ingredients. 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 is a deterrent to patentees who find that their invention is more rapidly made obsolete by the improvement and eliminate their monopoly profits. 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. The patent authority, then, needs to provide specialized incentives to patentee in fertile research areas. This takes the form of broader patents. However, 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.

Given that current patent law does not explicitly allow for multiple widths, inventors may seek ways to acquire extra breadth when they need it. One commonly discussed practice is the patent wall. An inventor discovers a particularly useful innovation, which is likely to be made obsolete rapidly. Before going to the patent office, the researcher engages in socially wasteful spending in order to generate several other nearly duplicate inventions. Patenting a variety of inventions in order to build this wall is supposed to be a way of acquiring breadth, since a subsequent innovation will only be a competitor if it infringes on *none* of the many patents in the wall. Potential infringers can respond by spending resources making their improvement appear to be non-infringing ("designing around").

A model of optimal patent walls is formulated as an extension to the sequential model introduced in section 4. More profitable innovations and innovations more prone to obsolescence attract walls. Since walls and designing around is wasteful, offering broader patents to the innovators who would waste effort building walls can be another desirable feature of a multiple-breadth patent menu. Indeed, the fact that it may be optimal to offer a variety of breadths is confirmed in the model with endogenous breadth through walls and patenting around.

Much existing work on optimal patent breadth deals with models of horizontal differentiation. The model of sections 3 and 4 is one of vertical differentiation. To see how the same concept of using menus of patent breadths applies to these commonly discussed cases, section 5 shows how unobserved heterogeneity can lead to the desirability of offering a variety of patent breadths in the horizontal differentiation model of Klemperer (1990). It is also noted that even if breadth is constrained to take a single value, the optimal patent may differ from the one that Klemperer (1990) finds in the presence of complete information.

As a patent authority, setting breadth may be difficult, in practice, due to the wide variety of products that are patented under a single statute. This is an issue in standard models of breadth such as Gilbert and Shapiro (199) and Klemperer (1990), but is even more apparent when the optimal patent calls for a variety of breadths being delineated. Overcoming this problem is the focus of section 6. Two approaches are considered. The first is simply to explain how current case law on patent infringement provides useful instruments for offering differing patent breadth. Doctrines imposed by the courts and now a part of the effective law could be explicitly written in or out of the various types of patents offered.

A second approach is to use fees paid by infringers to have the right to infringe. Such an instrument is explored within the vertical differentiation model, and it is shown that appropriate fees paid by infringers can implement a specific breadth of protection. Patents could be written with variable fees in order to provide differing degrees of breadth.

Section 7 attacks the issue of whether or not the problems raised in the theory are relevant to patentees. First, we look for evidence of patent obsolescence in the US patent data. The proposal of offering differential breadth is most easily justified when obsolescence is important, so socially desirable projects may not value additional statutory patent length. We measure obsolescence by looking at the effect of citations. First we look at patent renewal. We take recent citations (2 years prior to renewal and sooner) as evidence of no obsolescence, and find that recent citations contribute positively to renewal after 12 years, suggesting that some patents that are no longer cited have lost value. We compare claims directed at a given patent to claims directed at a patent that cites the original patent (indirect citation). Patents which are still being cited indirectly during their statutory term often cease to be cited directly, evidence that improvements can lead to patents losing their value.

Finally, anecdotal evidence is presented from interviews with practitioners involved in making patenting and research decisions. They indicate that the sort of problems that, according to the theory, may arise under a single patent breadth are in fact a concern. They also confirm that patentees seek additional breadth through patent walls.

2. The General Design Problem

Consider a patent authority faced with the following problem. An idea $\theta \in \Theta$ arrives randomly according to $G(\theta)$ to an innovator at time zero. 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 $\Pi(B, T, \theta)$.[1] It is continuous as well as increasing in the first two arguments. Property rights (B and T) are essential if the innovator is to make profits: $\Pi(0, T, \theta) = \Pi(B, 0, \theta) = 0$. The authority may charge a fee F for

---

[1] Gilbert and Shapiro specialize to the case where profits are $\int_0^T \pi(B, \theta)$.

the 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.

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 non-distortive 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 does not effect future innovators. Assuming that there is a better instrument than this sale of monopoly power is enough to maintain the results.

The patent authority solves the problem

$$\begin{aligned} & \max_{B(\theta),T(\theta),F(\theta)} \int_{\Theta} S(B(\theta),T(\theta),\theta)dG(\theta) \\ & s.t. \\ IR &: \Pi(B(\theta),T(\theta),\theta) - c - F(\theta) \geq 0 \\ IC &: \Pi(B(\theta),T(\theta),\theta) - c - F(\theta) \geq \Pi(B(\hat{\theta}),T(\hat{\theta}),\theta) - c - F(\hat{\theta}) \ \forall \ \hat{\theta} \\ Moral\ Hazard &: F(\theta) \geq 0 \ \forall \ \theta \end{aligned}$$

The problem is formulated as an optimal revelation mechanism. The patent authority sets a menu of patents, a breadth, length, and fee for each type θ. We will assume that all useful innovations are worthy of being implemented; the first constraint (individual rationality) says that the menu grants to each type sufficient profits to cover costs. The second constraint (incentive compatibility) ensures that each type θ accurately "reports" their type by choosing the appropriate menu item. The final constraint, moral hazard, ensures that worthless innovations are not patented. Since research costs are positive, some property rights will be conveyed.

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. 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.

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.

Consider the case where the following sorting conditions hold:[2]

Sorting condition:
$\Pi_{13}(B, T, \theta) > 0 > \Pi_{23}(B, T, \theta)$
$\Pi(B, T, \theta)$ is monotonic in θ

The type space θ is taken to index projects by there profitability, either up or down. High θ types value breadth more and time less. This sorting condition allows us to prove the following:[3]

Proposition 2.1. *Let $\Theta = \{1, ..., J\}$ be a finite set of types. 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 conferred. On the other hand, they value increments of time less because they are more likely to lose their market power before T due to non-infringing competitors. Their effective patent length is likely to be limited by the early arrival of competitors outside of their patent's breadth.

Currently, there is a sort of patent menu: most patent systems, including, now, the US, 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

---

[2] Subscripts denote derivatives or differences.

[3] The proof of the Propositions in this section is contained in an appendix.

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 2.1 implies that this is not optimal. However, anytime breadth is both costly (S strictly decreasing in B) and a viable instrument (Π strictly increasing in B), it is never optimal to charge fees and maintain a constant breadth across types.

Proposition 2.2. *Let* Θ *be a countable set of types. If* Π *is strictly increasing in* B *and* S *is strictly decreasing in* B*, then a patent of the form* $(B, T(\theta), F(\theta))$ *with* $F(\theta) > 0$ *for all* θ *cannot be optimal.*

Breadth is less costly than fees, since it allows time of protection to be reduced, whereas using fees require time to be increased to make up for the innovator's cost of paying the fee. The patent authority should always use the breadth instrument if it is available.

3. Breadth and Innovation Fertility: A Sequential Model

The model above takes as given the profit function of the potential patentee, in the spirit of work by Gilbert and Shapiro. Given the general results above, it is useful to consider a more structural model, describing the patent problem with preferences and technologies. We consider the case of an innovation which arrives and then may be superseded, so that the patent's effective length (the time until a non-infringing improvement drives the original innovator out of the market) is different from its statutory life.

3.1. Environment

Consider a vertically differentiated quality ladder. Before the first innovation arrives, a technology producing a good of quality normalized to 0 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, it's quality *improvement* over π is given by Δ, which is distributed according to a continuously differentiable cumulative density $F(\Delta)$. 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.

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.

In general, 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 bargaining 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 a patent provides:

given the setup, the patentee extracts all of the surplus from the consumer. We investigate cases where this does not hold later, and confirm that it is not crucial to the results.

It is assumed that the second innovator has access to the same set of patents. In principle one would offer different patents to the second innovator, but the menu is maintained to keep things as simple as possible. The second innovator simply chooses the patent with the maximum time, since breadth has no benefit. Note again that, given the specification, there is no cost to granting a patent to the second innovator.

3.2. 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 trade off of length for breadth. Second, it makes the solution to the design problem simpler to characterize and more intuitive, since it has only two instruments, effectively. We then show circumstances under which the menu is non-trivial, i.e. it is not only incentive feasible to trade B for T in the optimal mechanism, but it is also optimal.

The patentee makes profits π until a second innovation arrives and the innovation is bigger than B, until the patent expires. 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, and let the future be discounted by ρ. Discounted expected profits for the patentee are

$$\Pi(B, T, \theta) = \int_0^T e^{-\rho t}\pi[(1 - P(t, \theta)) + P(t, \theta)F(B)]dt \tag{3.1}$$

Notice that

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

since $P_2 > 0$, higher θ implies more likely arrival.

The derivative of the integrand in (**??**) with respect to B is

$$e^{-\rho t}\pi[P(t, \theta)f(B)]$$

where f is the pdf of F. Taking the derivative of that with respect to θ yields

$$e^{-\rho t}\pi[P_2(t, \theta)f(B)] > 0$$

which implies the cross derivative $\frac{d^2}{dBd\theta}$ of the integrand is positive. As a result, $\Pi_{13}(B,T,\theta) > 0$. The problem simplifies to the form of the previous section, from which we conclude that optimal fees are zero.

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

$$S(B,T,\theta) = \left\{ \begin{array}{c} \pi \int_0^\infty e^{-\rho t} dt + \\ E(\Delta|\Delta > B)(1 - F(B)) \int_0^T e^{-\rho t} P(t,\theta)dt + \\ E(\Delta) \int_T^\infty e^{-\rho t} P(t,\theta)dt \end{array} \right\} \tag{3.2}$$

The cost of patents is that they lead to slower arrival of second generation innovations. Society always enjoys π, either from profits or consumer surplus. The second term includes the incremental social gain from all arrivals up to time T which are implemented. Until T they are only implemented if they exceed B, in which case they provide an expected social gain equal to the expectation of Δ conditional on $\Delta > B$. This even has a probability of $1 - F(B)$, conditional on arrival. The last term includes the fact that all arrivals are implemented one way or another after T and generate incremental expected gain of the expectation of Δ from then on.

One way to better understand the preferences of society is to calculate the marginal disutility of each patent instrument. To calculate the derivative of S with respect to B, rewrite $E(\Delta|\Delta > B)(1 - F(B))$ in the second term of (**??**) as

$$E(\Delta|\Delta > B)(1 - F(B)) = \int_B^\infty \Delta dF(\Delta)$$

and compute the derivative $S_1(B,T,\theta)$, with, again, f denoting the probability density function of F:

$$S_1(B,T,\theta) = -Bf(B) \int_0^T e^{-\rho t} P(t,\theta)dt$$

The cost of breadth depends on the discounted probability of arrival and the value B and probability $f(B)$ of additional second generation products delayed. It is interesting to note that although high θ implies a high valuation of breadth, it also means that breadth is relatively costly to society, in terms of lost social welfare:

$$S_{13}(B,T,\theta) = -Bf(B)\int_0^T e^{-\rho t}P_2(t,\theta)dt < 0$$

Since P_2 is positive, S is more sensitive to breadth as θ rises. Fertile areas give rise to more second generation innovations, and so precluding them is more costly to society. This force can tend to prevent society from offering greater breadth to more fertile inventions. Much of the remainder of this section will serve to show, however, that it does not necessarily eliminate offering greater breadth to high θ inventions.

In order to calculate he derivative of S with respect to T, it is useful to rewrite S as

$$S(B,T,\theta) = \left\{ \begin{array}{c} \pi \int_0^\infty e^{-\rho t}dt + \\ E(\Delta|\Delta > B)(1-F(B))\int_0^\infty e^{-\rho t}P(t,\theta)dt + \\ E(\Delta|\Delta < B)F(B)\int_T^\infty e^{-\rho t}P(t,\theta)dt \end{array} \right\}$$

so that it is straightforward to calculate the derivative of S with respect to T, S_2, as

$$S_2(B,T,\theta) = -F(B)E(\Delta|\Delta < B)e^{-\rho T}P(T,\theta) < 0$$

The cost of time depends on how many patents are effected ($F(B)$), their value ($E(\Delta|\Delta < B)$), and the probability of arrival before T, $P(T,\theta)$. Granting additional time is increasingly costly in θ:

$$S_{23}(B,T,\theta) = -F(B)E(\Delta|\Delta < B)e^{-\rho T}P_2(T,\theta) < 0$$

Notice that both B and T are more costly the higher is θ. Because of this, it is possible to have non-trivial patent menus, in the sense that the optimal incentive compatible menu does not set B and T the same for all θ. When society offers more breadth to a patentee, it can offer less length and still offer the same reward. The cost to society of patents depends on the distribution of Δ across the effected improvements $\Delta < B$; the patentee only cares about the probability with which B is exceeded.

The cost of patent power can be described by the function $-S(B,T,\theta)$ which is minimized by the patent authority. The findings for the sequential model are summarized in the following proposition.

Proposition 3.1. *The social cost of patents, $-S(B,T,\theta)$, are increasing in the first two arguments. The marginal cost of each patent instrument is increasing in*

θ. The marginal profit value of B is increasing in θ and the marginal profit value of T is decreasing in θ.

3.3. An Example

To illustrate the role and optimality of non-trivial patent menus, consider a simple case where $F(\Delta)$ is a point mass on some $\bar{\Delta}$. 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 0 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 of 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 trade-off faced by the patent office. It is forced to provide breadth $B_2 = \bar{\Delta}$ to induce innovation in very valuable, high fertility areas; however, this protection is very costly when $\bar{\Delta}$ is large, and so it is only used 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 $(\bar{\Delta}, T_2)$ 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 $(\bar{\Delta}, T_2)$. 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. Patent Walls and "Designing Around"

The discussion so far has focused on the normative: how should patents differ when inventions differ? The patent statute, however, does not provide for this in practice. How do innovators respond? One common story is that patentees can acquire breadth by taking an initial innovation and build a "patent wall" by patenting many similar products. The rationale is that holding many patents makes it more likely that any improvement on the true innovation will be found to infringe some part of the wall. The sequential model is well suited to analyzing the way innovators respond to the possibility of patent walls.

The other end of patent walls is what is called "designing around." Designing around is the idea that a substitute or improvement, which would be found to infringe, is suitably modified for the sole purpose of avoiding infringement of existing patents. Walls may discourage this behavior by making it more difficult and hence costly to redesign second generation products.

Both of these behaviors are relevant to the normative issues raised. To the extent that patent law does not adequately provide breadth to the most fertile innovations, it may lead to inefficiencies in the wasteful spending of research resources in designing other parts of the wall, even when they have no incremental value of their own. At the same time, designing around can be seen as a wasteful activity, also wasting research resources.

4.1. A Model of Walls and Designing Around

Walls and designing around are incorporated into the sequential model as a simple, reduced form addition to the model. Suppose that the first innovator, upon realizing θ, can choose not only whether or not to spend c, but also to spend some additional resources w to build a wall. When an improvement arrives, second generation innovators can spend resources a designing around the initial patent.

Given a statutory breadth B, an improvement Δ infringes if $\Delta < B+\Omega(w)-K(a)$, where both Ω and K are increasing, concave, and equal zero when evaluated at zero. Patent breadth is costly to acquire, and the marginal cost is also increasing. The interpretation is that it may not be difficult to invent a few "reinventions" that are patentable and help the patentee defend the innovation, but it becomes increasingly difficult to develop non-inventions that will be helpful in defending the patent. Likewise, the cost of patenting around is increasing in the amount done, and is also increasingly costly at the margin: it is very difficult to patent around when Δ is far from $B + \Omega(w)$.

Working backwards, given an effective breadth of protection $B + \Omega(w)$, the second generation innovator must design around for any improvement $\Delta < B + \Omega(w)$ in the amount $\bar{a}(B + \Omega(w))$ which solves

$$K(\bar{a}) = B + \Omega(w) - \Delta$$

For an arrival at time t, this is preferable to waiting for the original patent to expire if

$$\int_0^{\max T} e^{-\rho t}\Delta dt - \bar{a}(B + \Omega(w)) > \int_{T(\theta)-t}^{T(\theta)+\max T} e^{-\rho t}\Delta dt$$

or, if Δ exceeds

$$\Delta_L(B + \Omega(w), t) = \frac{\bar{a}(B + \Omega(w))}{\int_0^{\max T} e^{-\rho t} dt - \int_{T(\theta)-t}^{T(\theta)+\max T} e^{-\rho t} dt}$$

The solution to the patenting around problem, then, is

$$a^*(B + \Omega(w), t) = \begin{cases} \bar{a} \text{ if } \Delta \in (\Delta_L(B + \Omega(w), t), B + \Omega(w)) \\ 0 \text{ otherwise} \end{cases}$$

Very small improvements are not worth the expense of patenting around. Large improvements are noninfringing on their own.

The patent wall problem is

$$\max_w \int_0^T e^{-\rho t}\pi[(1 - P(t, \theta)) + P(t, \theta)F(\Delta_L(B + \Omega(w), t))]dt - w$$

Now the patentee ceases to make monopoly profits if the improvement is at least Δ_L, in which case the second product is developed through patenting around.

There are a variety of simple regularities about the process of patenting around and developing patent walls which are useful.

Proposition 4.1. *The cutoff project implemented,* $\Delta_L(\cdot, t)$, *is increasing in the first argument.*

This is simply a consequence of the fact that the higher is the effective breadth, the higher is the cost of designing around, and hence the larger is the improvement required to motivate the undertaking. The more effective breadth the patent has (defined by $B + \Omega(w)$), the more of an improvement the second innovator needs to find it worthwhile to develop and patent around (if necessary). Since effective breadth is less costly at the margin when B is higher, the patentee acquires more effective protection when B is higher.

Proposition 4.2. *The effective protection obtained by the patentee* $B+\Omega(w^*(B))$ *is increasing in* B*. Also, the optimal patent wall* w^* *an hence the effective protection is increasing in* θ *and* π*.*

Since Ω is concave, the marginal cost of additional effective breadth is decreasing in B for any $B + \Omega(w)$, and the marginal benefit of effective breadth is independent of breadth, so therefore more is acquired the higher is breadth. On the other hand, the marginal benefit of additional effective breadth is increasing in θ and π; concavity implies the marginal cost of breadth is increasing. Optimality of w requires that marginal cost equals marginal benefit, and so increasing in θ and π The effect of B on w^* is ambiguous, and depends on, among other things, $F(\Delta)$. The latter part of the proposition presents a straightforward but useful comparative static: more profitable (per period) projects and ones more susceptible to obsolescence are the ones where walls are the biggest.

4.2. Normative Implications

Patent menus can have another role in this environment. Patent walls and designing around are accomplished by resources better used, in the eyes of society, elsewhere, if incentives could be designed properly. It is straightforward to calculate, as in the case without walls or patenting around, that the sorting conditions hold, and therefore the optimal fees are zero.[4]

Proposition 4.3. $\Pi_{13} > 0$, $\Pi_{23} < 0$.

---

[4] The only fact to remember is that both effective breadth and, therefore, the cutoff project implemented are increasing in the relevant arguments, as discussed above.

Once again, if it is useful to offer differing patents do different projects, then sorting should be done without fees. In addition to the reasons before, menus can serve to provide protection that would have been acquired at social cost.

5. Other Monopoly Costs of Patents

So far, the monopoly cost that has been the focus is the fact that patents may preclude second-generation innovations. The use of patent menus is not limited to these cases. Other, perhaps more conventional costs of monopoly can also be considered, and patent menus can be optimal in those cases as well. Such a possibility is demonstrated using a standard horizontal-differentiation model of patent breadth.

Consider the model used by Klemperer (1990). 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 $\bar{p}$ and a cost of substitution τ per unit in the product space away in from the patented good. Each consumer chooses to consumer one unit or none. They may consume either the patented good, which has price p, or a competitively produced substitute B units away, priced at marginal cost m, and so has effective price $m + \tau B$.

In the sequential innovations model, θ indexed the speed of new arrivals. Here θ will effect the cost of the competitor. Suppose that none of the products are possible before the arrival of the potentially patented invention. After the arrival, if the innovator chooses to invest c, the good which 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 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, thought, 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 in order 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, they all either substitute or do not consume.

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.

5.1. 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 (except for the highest type, who will be on the IR constraint) endogenous. By offering a broader but more short-lived patent, society loses relative to a long lived patent which offers the same profits, but 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 the costs of research; many are rebated more, due to rents from their private information.[5]

Proposition 5.1. *Consider the horizontal-differentiation model, and let c be sufficiently small and only one type of patent (B,T) is offered. It is not necessarily true that the optimal patent has $T=\infty$.*

More generally than simply in this model, unobserved heterogeneity impacts the optimal patent. This is something to keep in mind anytime patent length and breadth are chosen.

---

[5] Once again, the proofs of results in this section are contained in an appendix.

5.2. Patent Menus

Anytime the constraint binds for type θ and a given B, it also binds for a type $\theta' > \theta$. Clearly, then, $\pi_2 < 0$, and so $\Pi_{23} = \pi_3(B,\theta)e^{-\rho T} < 0$. The value of B is the shadow value of increasing the constraint at the rate τ. If the constraint does not bind, B has no marginal value to the patent holder. To calculate Π_{13}, note the first order condition from the patentees problem when the constraint binds

$$D(p) + pd(p) = \lambda$$

The Lagrange multiplier λ is the value of relaxing the constraint, and $d(p)$ is the pdf of D. Since $p = m(\theta) + \tau B$ at the constraint, taking the cross derivative of λ with respect to B and θ gives

$$\tau m'(\theta)\left(3d'(p) + pd''(p)\right)$$

where primes denote derivatives. This is positive if the last term is negative, under which $\Pi_{13} > 0$.

Under some restriction, then, the sorting condition apply. That is not, however, evidence that it is societally preferable to offer a variety of patents. Nonetheless, it can be. Consider a case with a finite number of types, 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. For instance, consider a case (proven to exist by the previous proposition), where an infinitely lived patent is not optimal if only one patent is offered. Begin with some patent $(\bar{B}, \bar{T})$, offered to all types. Consider, for the type with the lowest θ, denoted θ_L (i.e. the least flexible invention, the one with the least competition and the most profitability), offering a patent with infinite length that provides the same profits as alternative on the menu, i.e. $(\tilde{B}, \infty)$, where $\tilde{B}$ solves

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

Under the sorting condition, this is incentive compatible, since replacing B with T is most desirable for θ_L. But by Klemperer's proposition 2, conditional on a given reward, it is optimal to offer T as high as possible, and so offering the menu is beneficial.

This section, then, establishes two points: first, even without a menu of patents, the existence of heterogeneity effects 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.

6. Implementing Patent Breadth

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 difficulty. Here we address that issue by discussing operational ways to implement B. The first set of these pertains to employing the many ways patent breadth is currently defined. Another perhaps more novel approach is to consider the possibility of allowing infringement in exchange for a predetermined buyout fee. The impact of buyout fees is examined within the framework of the sequential model introduced earlier.

6.1. Currently Available Breadth Instruments

The patent statute is relatively vague in its definition of what a patent protects against. Taken at its most literally, a patent is a set of claims, listed at the application date, which define the innovation. The actual scope of the protection extends beyond these claims, however, for an obvious reason: noninfringing improvements are meant to be substantively different from the patented innovation, not just descriptively different. As a result, the case law and the treatment by the patent office has included a variety of extra protections upheld by the courts.[6]

6.1.1. What can be claimed?

One crucial element of patent breadth is the specificity of the claim. Patent law states that all claims must disclose enough information for "any person skilled in the art to make and use" the innovation. In fact, recent Patent Office rulings have extended patent scope. One example is the case of Genentech.[7] Scientists produced two human proteins using a given process. The process was not new, but the scientists were the first to use it to produce human proteins. Genentech claimed rights to the principle of using this method to generate human proteins. They were allowed to, in essence, patent the principle.

---

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

[7] Once again, see Merges and Nelson (1994) for a more detailed account.

6.1.2. The Doctrine of Equivalents

Given a set of claims, the breadth of protection is not necessarily limited to the claims themselves. 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 where the Supreme Court upheld the Doctrine, *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.[8] Warner Jenkinson filed for a patent a year later, claiming a technique suitable for a pH of five. Hilton Davis claimed that their patent should extend as low as five by the Doctrine of Equivalents, despite the fact that (pH is a logarithmic scale) their claimed process was ten times as basic as the Hilton Davis procedure.

6.1.3. When are claims made?

In some cases, not all the capabilities of a product are known at the time of the discovery, and so some are left out of the patent's claims. However, when suing for infringement, a patentee can claim infringements on grounds that all of the capabilities of the infringing discovery could have been accomplished by the original patented product. For instance, one of the arguments used by Hilton Davis was that despite the claim of pH levels from six to nine, their process could have succeeded at a pH of 5. Given the current law, it is legitimate to make these kinds of ex-post claims in a patent infringement case.

6.1.4. Incorporation of these doctrines

In its recent ruling in the *Warner Jenkinson v. Hilton Davis* case, 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

---

[8] Interestingly, Hilton Davis originally filed for a patent where the claim simply said that their process worked for pH's less than nine. The patent examiner required that the claim be made more specific.

now an element of the law which 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.

Other changes and evolutions of patent law could be laid out in a patent statute that allowed for multiple types of patents. Can a patentee reserve the right to add future claims, if they are discovered to be possible under the original invention? Can the "principle" of the invention be patented, or only the specific accomplishments? In each case, the inclusion of the protection as an option on the patent menu would be clearly definable.

A menu of patents is not novel in the US, in a sense. The US has incorporated renewal fees, so that patent of different lengths have different fees. The fees, however, are quite small. In terms of breadth, the US does have one sort of patent menu. Asexually reproduced plants may be patented under either of two different patents. First is the standard utility patent. Second is a plant patent, which provides protection if and only if the infringer is grafted from a patented plant.[9] The function or structure of the plant is irrelevant under a plant patent, unlike a common utility patent, meaning a different set of subsequent plants is protected.

6.2. Buyout Fees

Given that the discussion above only focuses on a few ways to define breadth, and since the optimal patent menu may involve patent breadths not achievable with existing definitions of breadth, another way in which breadth can be implemented is useful. If improvements must pay a fee for each infringement, regardless of the level of improvement, small improvements may be avoidable due to the small gain they offer an infringing innovator. These fees are termed "buyout" fees, to distinguish them from the F introduced above.

In practice, the court system is used to determine infringement. One way to think of this process is in the language of the sequential model: the courts determine if Δ is greater than B. Because this process is so uncertain, it seems reasonable to assume that not only θ but also Δ may be unobserved. Another way to say this is that it may be difficult, given that patents are arbitrated in court and come in many sorts, to clearly define B in the statute. We will assume

---

[9] >From the Franklin Pierce Law Center web page, http://www.fplc.edu/TFIELD/plfip/plfipCom.htm.

that by the time that the second innovation arrives, π is observable, but θ and Δ are not. The history of the patented good, then, is sufficient to know it's value, but the new product does not have the same record.

6.2.1. Fees Paid to the Government

Instead of defining a patent breadth, the government policy can take the form of a fee charged to the second innovator, paid to the government, in exchange for freedom from the initial patent. A patentee who chooses a high fee is implicitly choosing a broad patent, since only large Δ projects will be sufficiently profitable to make paying the fee worthwhile for the second innovator.

Denote the buyout fee by $\phi_t(B)$, and define it by

$$\phi_t(B) = (1 - e^{-\rho((\max T)-t)}) \int_0^{\max T} e^{-\rho t} B dt$$

The buyout fee ϕ is exactly equal to the gain in gross profits from immediate production of an innovator with improvement B, who chooses in this model simply the maximum time of protection. Since profits are increasing in Δ, it is straightforward to see that only projects $\Delta > B$ are implemented, since only then is it worthwhile to pay the fee.

Implicitly, this patenting problem has had at its heart an inefficiency in the transactions made between the first and second innovator. Because of this, some useful projects are delayed by patent protection. The essence of the buyout system is that, in such a circumstance, the fee serves as a sort of pre-committed contract, where the initial patentee is forced to "sell" the patent right for ϕ. This definition of B makes it straightforward to implement breadth versus length patent menus which are often optimal. Since the original patent still runs until an improvement of B is reached, the profit function for the innovator is unchanged and the sorting conditions hold. Moreover, S is unchanged, so all of the normative implications remain. Menus of breadth and length, in this case in the form $(\phi(B(\theta)), T(\theta))$, may be useful.

The buyout fee is a different fee from the one (F) described in the earlier section. Intuitively, though, the first proposition of the paper pointed out that if government revenue is better raised through sources other than monopoly power, it is preferable to find patent instruments other than fees. One natural alternative is to have ϕ paid to the first innovator as a reward for the useful product.

A patent policy based on a mandated buyout fee is a sort of pre-commit licensing agreement, which can be valuable when ex post licensing leads to inefficiency.

Buyout fees might be used in themselves as a simple-to-define breadth instrument, or in conjunction with the definitions of breadth discussed previously in order to provide a variety of well defined patent breadths to choose from

7. Evidence on Obsolescence

The policy recommendation above, to sort across different products using different patent breadth, applies to the case where different innovations have different spillovers to other products, in terms of making improvements come along sooner. Especially useful innovations may require a great about of breadth that is not optimal to provide to other projects. The crucial feature is that more socially desirable patents may be less profitable due to their value as building blocks to future innovations.

7.1. Evidence from US Patent Data

The sample used for this analysis consists of all utility patents that were granted in 1983. The citation history of these patents was constructed taking all citations up to 1995. The choice of sample was motivated by two factors: 1) To have long citation histories; 2) Including patents for which renewal fees, instituted at the end of 1979, apply. The number of patents in this set is approximately 54,000, generating as of 1995 almost 315,000 citations, approximately 5.7 cites per patent.

One feature of the model above is that some projects are made obsolete before their patent term expires. This is what leads some firms, predicting rapid improvements leading to obsolescence, to prefer broader patent protection. Using US patent data, we find evidence that obsolescence is important. Since the patent data provides no direct evidence on value, we take the common stance of using patent renewals as a measure of value.[10] To measure obsolescence, we use patent citations. We view recent citations to a patent as evidence that the patent is still relevant, and hence generating value. We perform the following experiment. We regress, as a logit model of probability of renewal, the effect of both total citations and recent citations (citations made within 2 years of the renewal date). We include total citations since the current literature has shown (Trajtenberg (1990)) that citations are a positive predictor of a patents value. We also include number of claims in the patent as a way of holding the breadth of the patent fixed;

---

[10] For instance, Pakes and Schankerman (1986) and Pakes (1986) use renewal data from European patents to estimate patent value.

however, claims are never significant. For the renewal after 4 and 8 years, we find that recent citations has a positive but insignificant effect on renewal. We take this as evidence that most patents have at least a few years of effective life. At the 12 year renewal, however, we find that not only are total citations significant contributors to the probability of renewal, but that recent citations also predict positively and significantly. The results of that logit estimation are present below.

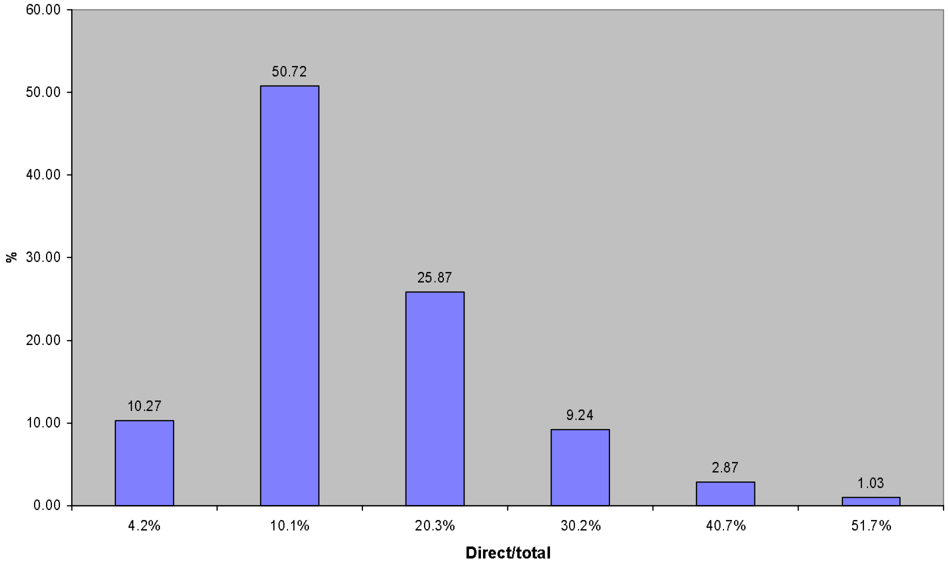
Variable	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCEPT	0.6371	0.0193	1086.5391	0.0001		
CLAIMS	0.00194	0.00128	2.3228	0.1275	0.009927	1.002
CITE12	0.0313	0.00263	141.6545	0.0001	0.121063	1.032
CITE10_12	0.0342	0.0106	10.4540	0.0012	0.033266	1.035

The odds ratio gives the additional probability of renewal for an increment to the variable; that is, a citation in the first 12 years increases the probability of renewal by 3.2%, but one in the two years prior has an additional 3.5% increase in the probability of renewal. Recent renewals are almost twice as important as other renewals.

We also explore the protection a patent provides as its statutory term proceeds. Citations are decomposed in the following way. Consider a patent p. Let D(p) denote the set of all patents that cite p during our sample period. We use the letter D to denote direct citations. Let I(p)=D(D(p)), denote indirect or second round citations for this patent. This set consists of all patents that cite any patent that directly cites p. Finally, let C(p)=I(p) U D(p) denote the set of direct and indirect citations for this patent. The following statistics were computed with a 5% random sample of the population of all utility patents granted in 1983 with between 10 and 25 total citations by 1995. This selection is made to focus on "successful" patents which generated a significant number of citations.

Figure 1 gives a frequency distribution of the ratio of direct to total citations. Higher values for this statistic reflect a patent that is not "forgotten" easily. Indeed, if all citations to downstream patents cited the upstream ones too, then the ratio would be one.

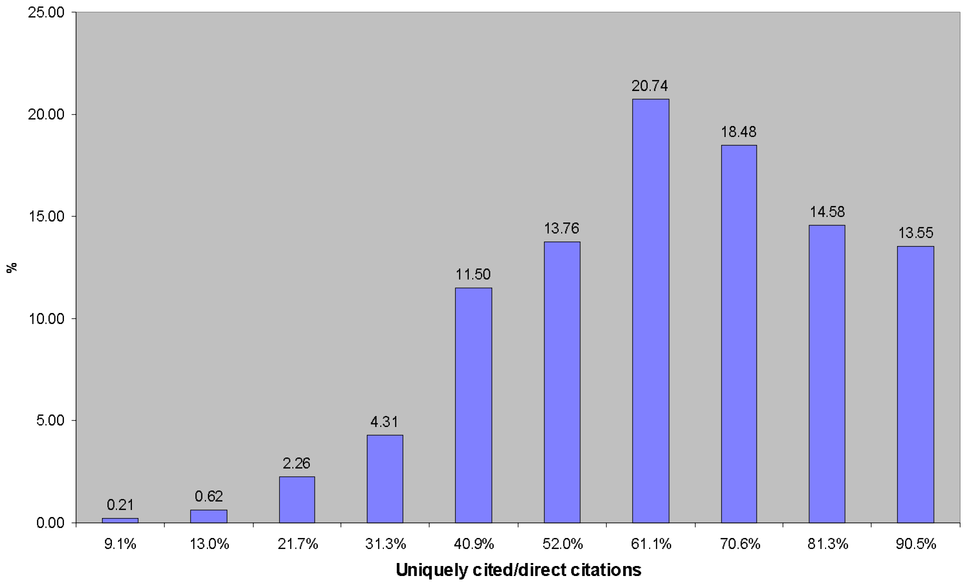
Figure 1. Direct/Total Citations



The mode and median are at around 10%, and there is quite a large variation. This low number may thus reflect that patents in C(p) do indeed substitute the original patent, receiving with exclusivity most of the new citations.. Notice that the patents we consider are fairly successful ones, as we restrict our sample to those that have between 10 and 25 citations.

Figure 2 provides a frequency decomposition of direct citations by exclusivity. For a given patent p we say patent q in C(p) is an exclusive citation if it does not cite other patents that directly cite p. Thus, the set of exclusive cites of p are all those patents in C(p) that do not belong to C(C(p)). A low value of C(p) could be interpreted as conferring more ambiguous property rights, since in that case a large proportion of those patents that cite the original one also cite one in its downstream.

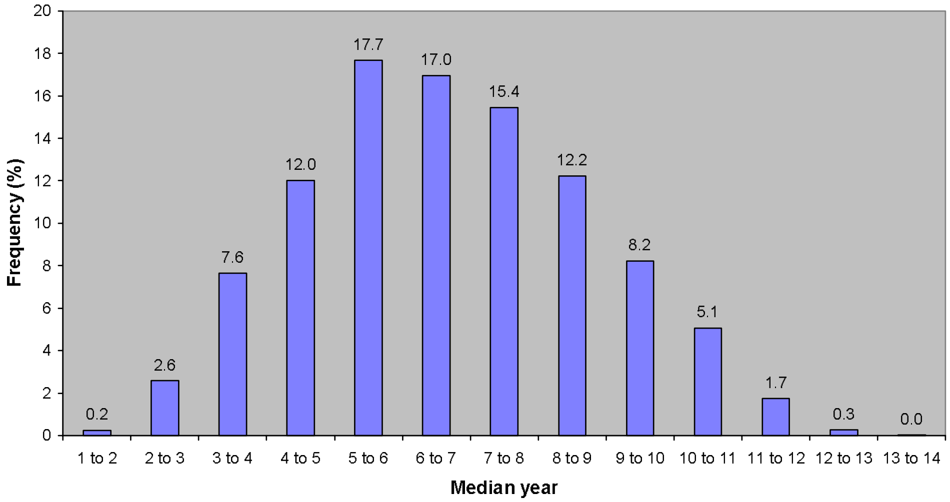
Figure 2. Uniquely Citing



The median and the mode are around 60%, indicating a relatively high degree of non-competing citations, or lack of complementarity.

Figure 3 gives a frequency distribution of the median citation year for a subset of patents. This frequency distribution is based on all utility patents granted in 1983 that had from 10 to 25 citations in the period 1983-95, almost 8,000 patents. For each patent, we computed the median citation lag. The figure provides a frequency distribution of these medians. It is important to notice that due to the existence of an end period, there is right truncation at 14 years. This certainly contributes to the decrease in citation frequencies exhibited after year 6.

Figure 3. Median citation years



This figure shows, quite strikingly, the existing dispersion in patent diffusion curves, as measured by citations. It also suggests that citation lags tend to be concentrated around 6 years

The summary statistics presented here, though quite preliminary and rudimentary, suggest that the design issues considered in this paper can be of considerable significance. One potential problem is that some highly useful, fast to obsolete projects may not be undertaken given the current patent offered in the US. To understand this part of the role of patents , we discussed the innovation process with some practitioners in research and development who make patenting and investment decisions.

7.2. How do innovators react to current policy?

In order to get a better handle on the way current policy impacts research and development, we interviewed several people involved in various areas of research. Fundamentally, the issue presented here is the question of whether socially more valuable projects may be privately less valuable when profits are eroded from socially beneficial future innovations, and whether current policy leads to the failure of researchers to invest in these projects. In total, we talked with individuals at large corporations such as Kodak, Xerox and Corning, as well as a research chemist at Aspen Research, a contract research company.

It was clear from our discussions that the availability of protection does effect the projects firms undertake. One interesting case of this came from talking to a contract research firm. The firm does a variety of problem-solving research for a variety of clients, including developing efficient quality control standards and analysis of product failures. Often, this research leads their scientists to contemplate patentable innovations, often with broader application, which might be especially useful to the client, but not part of the client's original order. Their procedure often, then involves contacting the client and discussing whether or not further research on the idea is something that the client would like to fund. In some ways, this is the story of an idea, the intrinsic value of which seems high, which requires funding. In conversation with this firm we asked what happens after that. The response was interesting. They said that their procedure is to approach the firm about funding the project. The agreement typically involves all or some of the rights to the patent to the client funding the project. They said that it was not uncommon for the client to agree with the scientists about the potential usefulness of the invention, but for the invention not to go forward due to problems with limited ability to protect the patent from innovators.

Another manager listed their criterion for patenting. Number one, of course, was value to the firm; number two was potential value to outsiders. This firm is very concerned with the negative consequences to its own market position of future inventions made by competitors using the knowledge embodied in the original patent. In some cases, an answer may be found in avoiding patents in favor of trade secrecy; sometimes, though, when the final product is easily reverse engineered, even this road is not possible. Trade secrecy can have cost, naturally, associated with limiting the transmission of information, with complete avoidance being an even more societally costly alternative.

A similar story was told by a manager in a large company with its own research laboratory. In their case, ideas flow to researchers in the laboratory, who then pitch the ideas to a set of people charged with developing the corporation's strategy toward the innovation. The researcher is primarily charged with coming up with the most useful inventions possible. Then the innovation is evaluated for profit potential. When asked about the latter part, it seemed clear that current patent breadth can lead to the avoidance of certain projects. In particular, the manager discussed a particular rival who has taken up the strategy of eliminating research and imitating their developments. The presence of this rival, in particular, means that some innovations that seem to have the potential for significant value end up receiving no support when they are vulnerable to encroachment by the imitator.

Interestingly, it was pointed out that the research cost c should sometimes be interpreted broadly. In cases where it is determined that the patent protection will not make a project profitable, often the project is already to a stage of patent readiness. The question is whether or not to pay the substantial costs of "commercialization," (a term of a practitioner for the investment made at this point), knowing that much of that cost generates benefits for both the firm and its competitors. A patent is unlikely to lead to anyone producing without the commercialization costs being paid, but paying the commercialization costs may not be worthwhile if the patent is not broad enough, to the extent that these investments benefit all potential market participants.

The notion that some innovations would gladly trade broader protection for less time protected was well received; it seemed, as well, that the sorts of projects for which this would be most welcome are, naturally, the ones where the inherent value seems high but the appropriable profits are low.

That innovators respond to current policy by building walls is not a novel finding, but it was confirmed here. It was, in virtually all cases, the practice of the company to develop a patent strategy toward an innovation that frequently involved resources to be spent on other patents whose primary purpose is to defend the initial one. "Broadly applicable" ideas are the ones where this is most common, according to practitioners; the company holds off on the initial patent until it has a clear patent for many of the foreseeable applications. This is evidence that firms are not able to appropriate their value as building blocks to future patents which incorporate their ideas. Instead, they work to patent many possible applications rather than delegate these tasks to the most productive firm. In addition, delay of the initial patent has natural costs of delayed information transmission, as well.

This is related to the idea of being able to patent a "concept" as discussed in the section on patent law above. Such broad protection can be dangerous, but in many areas, such as some of the inventions we discussed with the practitioners, it may be necessary. This paper suggests one way to deal with this is to provide such protection only for a few years.

8. Summary

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 which 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, as well as cases where extra patent protection can be acquired by innovators and destroyed by competitors at a cost.

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 must rely on a revelation of types by the patentees. To the extent that breadth can be used as an instrument, it is preferable to sorting with a patent fee. In almost all cases, it should be used as part of the menu. There are a variety of ways to define patent breadth, using an explicit menu that combines the case law established for deciding patent infringement cases and buyout fees paid by infringers.

There is some anecdotal that the heterogeneity of projects forces innovators to consider exactly the sort of questions the theory suggests a single type of patent protection raises. Some ideas are simply too useful to competitors to be worth the substantial costs of research. The profitability of a project given a type of protection is not necessarily increasing in the total value of the project.

US patent data provides evidence that patents are often superseded and their value lost prior to the end of their statutory terms. This is consistent with the model presented, which leads to the value of considering a policy of offering multiple patent breadths. Understanding the heterogeneity in patent fertility is an important question if such a policy is to be considered further.

A final point is that unobservable heterogeneity can effect optimal patents even when one restricts attention to offering a single patent type. A patent which 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. It may be misleading to design a patent for one type of innovation in isolation without considering the multitude of different innovations which will be protected by that same patent policy. At a minimum, this sort of heterogeneity should be considered when designing patents. One potential way to deal with this issue is to use a menu of patent breadths to sort different types of innovations; more research into the potential gains from such a policy seem worthwhile.

References

[1] Cornelli, F. and M. Schankerman. "Optimal Patent Renewals." Forthcoming, *The RAND Journal of Economics.*

[2] Gallini, Nancy. "Patent Policy and Costly Imitation." *The RAND Journal of Economics*, Vol. 23 (1992), pp. 51-63.

[3] Gilbert, R. and Shapiro, C. "Optimal Patent Length and Breadth." *The RAND Journal of Economics*, Vol. 21 (1990), pp. 106-112.

[4] Klemperer, P. "How Broad Should the Scope of Patent Protection Be?" *The RAND Journal of Economics*, Vol. 21 (1990), pp. 113-130.

[5] Merges, R.P. and Nelson, R.R. "On Limiting or Encouraging Rivalry in Technological Progress: The Effect of Patent Scope Decisions." Journal of Economic Behavior and Organization Vol. 25 (1994), pp.1-24

[6] O'Donoghue, T., Scotchmer, S., and Thisse, J-F.. "Patent Breadth, Patent Life and the Pace of Technological Improvement." Working Paper 95-242, Department of Economics, University of California, Berkeley.

[7] Pakes, A. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica* Vol. 64 (1986), pp. 755-784.

[8] Pakes, A. and M. Schankerman. " Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period." *Economic Journal*; Vol. 96 (1986), pp. 1052-76.

[9] Scotchmer, S. "On the Optimality of the Patent System," Working paper 236, Goldman School of Public Policy, University of California, Berkeley.

[10] Trajtenberg, M. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *The RAND Journal of Economics* Vol. 21 (1990), pp.172-187.

[11] Warner Jenkinson Company, Inc. v. Hilton Davis Chemical Co., Supreme Court of the United States No. 95-728.

A. Proofs

Proposition 2.1

Lemma A.1. *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 $l > j$ 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, but reducing patent length instead, 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. ■

Lemma A.2. *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 possiblities: 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 $l < j$ $(l > j)$ must strictly prefer the k-patent (l-patent) to the j-patent, so by lemma **??** 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. ■

Proof of Proposition 2.1.

Proof. For any $j = 1, ..., J$, let $B(j)$ 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 **??**, $B(A) \subset A$. But this implies that there is an incentive compatible patent menu where all fees in A can be reduced. Such patent menu could never be worse than the given one and by an appropriate reduction in breadth or length is potentially better. ■

Proof of Proposition 2.2.

Proof. Denote the uniform patent by (B, T, F). Consider the type for which Π_1 is minimized, and call it $\bar{\theta}$. For that type, offer an alternate patent $(B', T, 0)$, where $\Pi(B, T, \bar{\theta}) - F = \Pi(B', T, \bar{\theta})$, so $B' < B$. This patent is not desirable for any other type by the selection of $\bar{\theta}$, and hence is IC, and changes no ones payoff, and hence is IR. But since $B' < B$, it offers higher social value for the case of $\bar{\theta}$, and is hence preferable to society. ■

Proof or Proposition 5.1

Proof. 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 $p \leq m(\theta) + \tau B$ does not bind, i.e. the type θ_H patentee enjoys monopoly power forever under the proposed patent. If c is small enough, the costs associated with any patent (B, T) which 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_H's monopoly lasts, together with increasing B to satisfy the IR constraint of θ_L is always welfare improving. ■