

A Nonobvious Discussion about Patents ¹

Suzanne Scotchmer
scotch@berkeley.edu

Professor of Law, Economics and Public Policy, U.C., Berkeley

**Preliminary Draft:
Before quoting, check with author for updated version.**

draft August 28, 2008

¹This paper is a revision of my Frey Lecture at Duke University in April, 2008. It was prepared while I was Visiting Professor of Law at New York University. I thank the law school for its hospitality, and also acknowledge support of the Toulouse Network for Information Technology. I am grateful to Steve Maurer, Rochelle Dreyfuss, Mark Lemley, and Kathy Strandburg for detailed comments. I also thank Tony Reese and Arti Rai for discussion.

Abstract

Nonobviousness is surely the most vexed requirement for issuance of a patent. Several calls for patent reform have focused on tightening it up. However, I argue in this paper that the nonobviousness requirement, interpreted as the patentability standard, cannot be considered in isolation from breadth, which is another essential policy lever. The breadth of a patent establishes which rivals must license. I give an interpretive account of what we can learn from the economics literature on the optimal design of patentability and breadth standards. I argue that, where innovation is cumulative, breadth is as important a policy lever as the patentability standard. In the cumulative context, patent breadth can encourage investment in small improvements as well as large ones, without threatening profit.

JEL Classifications: O34, K00, L00

Keywords: patents, breadth, cumulative innovation, nonobviousness

1 Nonobviousness: Economic Concepts and Legal Concepts

Beginning in the industrial upheavals of the 19th century, there has been a vigorous policy debate as to whether patents are helpful or harmful in supporting innovation. The vigor of the debate waxes and wanes, but the nature of the arguments remains the same.¹ The gist is that, although patents are intended as an incentive, innovators can protect themselves in many other ways, and patents have the downside of stifling the use of knowledge, possibly to the detriment of further progress.

Even among those who admit the usefulness of patents, there is a question of how patents should be designed.² The requirements for obtaining a patent are utility, novelty, and nonobviousness, as well as appropriate subject matter. In addition, the disclosure in the patent must be enabling to other practitioners.³ These requirements, especially nonobviousness, leave considerable room for interpretation. With nonobviousness in particular, it has been hard to go from an intuitive, sensible meaning of the word to a legal doctrine that serves the incentive goals of the patent system. Commentators and even the Supreme Court are looking to nonobviousness as a hook that will allow us to reel in a patent system that has overreached.⁴

Since economists are interested in incentives, they stylize the patent debate with policy levers that bear directly on incentives, and do not worry too much about their legal basis. The ones I focus on here are the “patentability” and “breadth” standards. The interpretation of the legal rules, especially nonobviousness and enablement (and written description), has implications for both.

It is in the claims of the patent that legal doctrine and the economic policy levers come together. The *prima facie* test for patentability is whether the claim would have been obvious in light of prior art, presumably on the circular theory that, if it is obvious, it is not an invention.⁵ Echoing case law, the United States Patent and Trademark Office publishes Guidelines for how to apply this test. They include, for example, commercial success, long felt but unresolved needs, failure of others, and the rapidity with which innovations are made.⁶ An economic interpretation is that many of

¹See, for example, the fascinating account of Machlup and Penrose (1951), published in the run-up to the patent reform of that decade. In *A Patent System for the 21st Century*, 2003, National Academies of Sciences, a panel of experts brings this debate to modern technologies, in light of 20th century scholarship on the subject, not focused so much on whether we should have a patent system, but on how it should be designed.

²A large part of my 2004 book *Innovation and Incentives* concerns the incentive effects of various design elements. For legal recommendations, see, again, *A Patent System for the 21st Century*, 2003, National Academies of Sciences.

³35 U.S.C. Sec 103.

⁴See, again, *A Patent System for the 21st Century*. See also the recent Supreme Court decision, *KSR v. Teleflex*.

⁵Section 103(a) of 35 U.S.C. reads in part “A patent may not be obtained ... if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art.”

⁶See the USPTO’s *Manual of Patent Examining Procedure*, section 2141. Many of the tests for nonobviousness that are outlined there codify the secondary considerations for nonobviousness laid out in *Graham v. Deere* 383 U.S. 1 (1966).

these considerations are proxies for cost.⁷

As to breadth, the applicant can at most claim what she has enabled and described in the patent application. Enablement and the written description do not seem to leave much room for policy making. However, in *Graver Tank*, the U.S. Supreme Court extended exclusivity to any device that “performs substantially the same function in substantially the same way to obtain the same result” as the one claimed. The resulting doctrine is called the doctrine of equivalents, and it opens the door for using the breadth of claims as a policy lever.⁸ Economists have various stylizations of breadth,⁹ but all of them describe how “close” other devices or products must be in order to infringe.¹⁰

Of the two economic policy levers, I will argue that much of the economic work is done by breadth, not by the patentability standard. In fact, where applied to a single innovation, the breadth standard and the patentability standard are two faces of the same exclusive right. The patentability standard sets a threshold, but breadth does most of the work in determining profitability and deadweight loss in the market.

This is illustrated by an example of Barton (2003), namely, the cardboard sleeves that are used on coffee cups to protect the fingers. Most of these sleeves display patent numbers, and Barton uses the example to suggest that patent law has gone astray. It is surely obvious, in a plain-text meaning of the word, that one can protect the fingers by wrapping a hot beverage container with cardboard.¹¹ However, this is an incomplete analysis. I agree that the patents do not satisfy a straight-face test for nonobviousness¹², but where is the harm? If the various patents on the cardboard sleeves are mutually noninfringing, the market is not substantially different than if no patents had issued at all.

⁷Merges (1993) stresses failure of others (or potential failure of others) when he suggests that R&D programs with uncertain outcomes should receive high reward in case of success. The expected cost of innovations, averaged over the successes, is higher when the probability of success is lower. Erkal and Scotchmer (2007) stress long-felt need when they argue that rewards should increase with the length of time that passes before a known market niche is filled. Their rationale concerns cost. They assume that ideas to serve the market niche arrive randomly, and eventually an idea will have low enough cost to elicit investment. There is a social tradeoff between filling the market niche and waiting for a lower-cost idea. As time passes, rational pessimism sets in. If it seems there will be a long wait for a low-cost idea, it is better to become less discriminating.

⁸Merges (1993) makes this point at page 18 by describing claims that relate to a chemical process. The claimed process occurs at certain concentrations. He argues that the concentrations that pass the nonobviousness hurdle can still be expanded as a matter of breadth, and that breadth is ultimately restrained by the enablement requirement.

⁹These are described in chapter 4 of my (2004) book, *Innovation and Incentives*.

¹⁰In this paper, I use the word “infringe” as a description of the product and the prior patent rather than an assertion that the product is actually being marketed without a license from the prior patentee. This is a shorthand way of saying that the product *would* be held infringing if it were marketed without a license.

¹¹However, the patent appellate court seems to have rejected plain-text meaning. In *In re Lee*, they opine that “[t]he ‘common knowledge and common sense’ on which the Board relied in rejecting Lee’s application are not the specialized knowledge and expertise contemplated by the Administrative Procedure Act. Conclusory statements such as those here provided do not fulfill the agency’s obligatio.” 277 F.3d 1338, 13rr (Fed. Cir. 2002).

¹²Can you read this patent with a straight face?

The roles of the patentability and breadth standards become more tangled when innovation is cumulative, because the breadth of *other* patents, together with the patentability standard and the breadth of the innovator's *own* patent, may present alternative ways to protect an innovation. There are various ways that this can happen, with similar, but not identical, economic effects. That is the main subject of this article.

By “cumulative” innovation, I mean that a patented technology creates value for follow-on innovators, either because the first technology is essential for the follow-on, or because it reduces the cost of the follow-on, or because it accelerates the next discovery.¹³ The incentive problems that arise with cumulateness are the main subject of this inquiry. My thesis is that, especially in the cumulative context, we cannot have a sensible discussion about nonobviousness, or how to reform the legal rules to apply it, without also discussing patent breadth.

The nonobviousness debate seems to conflate two questions. First, should exclusive rights ever be awarded for things the applicant did not invent? Second, should an inventor get an exclusive right to *everything* he invents?

The answer to the first question (no) seems uncontroversial. Giving rights on technology the applicant did not invent does not serve an incentive purpose. Ensuring that this does not happen is probably the main gatekeeping function of the nonobviousness standard. If a technology is already known, it is surely “obvious.”

But the second question raises what it means to invent. It is precisely in the cumulative context where this becomes murky. For example, if a prior innovator has lowered the costs of achieving a follow-on innovation, is the prior innovator an inventor of the follow-on? If a prior innovation is a necessary precursor, does that strengthen the claim of the prior innovator to be an inventor of the follow-on? Indeed, what does it mean for rights to be exclusive?

Patents are different from other intellectual property regimes in that the “exclusive right” to a technology can be possessed by more than one party.¹⁴ That is, patent law has the special feature that patents can be “blocking.”¹⁵ All owners of blocking patents have an equal ability to keep the blocked technology from coming to market; each has an exclusive right. This is where the multiple claims to be an inventor come together, and why multiple inventors divide the profit.¹⁶ The division of profit is governed by a limited set of policy levers, in particular the breadth and the patentability standards, both governed by claims and by the nonobviousness requirement. That is the subject of this paper.

¹³This is the definition in Scotchmer (1991). See also Merges and Nelson (1990).

¹⁴Need citation. Merges?

¹⁵“Blocking patents” is an informal term that refers to the circumstance that a technology is at once patented and infringing on another patent. Both patent holders can block the technology from being commercialized. Lemley 1997?

¹⁶See Green and Scotchmer (1995) discussing how the division of profit plays out in the case of two stages.

There are many other policy levers of patent law that I do not discuss except incidentally, such as research exemptions, rights regarding reverse engineering, an independent invention defense, and others. I focus on breadth and patentability because the nonobviousness requirement links them.

2 Breadth and Patentability: Substitutes? Complements?

A follow-on innovation can be protected either through the exclusivity provided by its own patent, or by an exclusive license on an earlier patent it infringes, depending on that patent's breadth. The profit is protected if the follow-on innovation is (i) not patentable but infringing on prior patents, (ii) patentable but not infringing on prior patents, or (iii) patentable and infringing on prior patents. All three schemes give an exclusive right to use the second innovation, either to the first innovator or to the second innovator or to both.

However, the three schemes are not equivalent. They generate different incentives for research because they change how the profit is divided. In case (i), where follow-on innovations are not patentable, the prior (infringed) patent holder is likely to get most of the profit. The infringed patent holder can either invest in follow-on products directly, or can make agreements with prospective second innovators who bring proposals. In case (ii), all the profit of the follow-on innovation goes to the follow-on innovator, who does not have to license. In case (iii), a prospective follow-on innovator will have to pay the first patent holder, but may get more profit than in case (i) because he has a better bargaining position. He has the option to invest without a prior license, and then negotiate ex post.¹⁷ Because he has a patent, he has an ex post hold-up value on his own invention.

Because the three schemes privilege different parties, they have different incentive effects, even if they are substitute policies to generate the same total profit. Which one is better depends on one's prediction about transactions. For transactions optimists, as Rochelle Dreyfuss has called them, case (i) is better than case (iii), which is better than case (ii). For transactions pessimists, case (iii) is better than case (ii), which is better than case (i).¹⁸

This story assumes there are just two stages of research. With two stages of research, there is a natural asymmetry between the first, or basic, innovator, and the follow-on innovator. The

¹⁷This raises the question of research exemptions. See Eisenberg (1988). If the follow-on innovator is using the knowledge disclosed in the prior patent, it is essentially a research exemption that makes this work. In the wake of *Duke v. Mady*, there is no longer a research exemption, and that should be taken into account, especially if a prohibition on experimentation is enforced.

¹⁸Transactions optimists believe that deals can take place before the investments in follow-on innovations occur. As a consequence, once the basic innovation has been achieved, the follow-on innovation is not in jeopardy, at least not in a way that can be remedied through breadth and patentability. (Green and Scotchmer, 1995) If the follow-on generates positive profit, the first innovator will make an ex ante deal to ensure that the investment takes place. Transactions pessimists worry that scheme (i) is not robust. If licensing agreements with the prior patent holder are unlikely to materialize before investment, the follow-on innovations might be stifled. Anticipating that, investment in the underlying basic innovations might also fail. See Lemley (1997) for a similar argument, also comparing copyrights to patents.

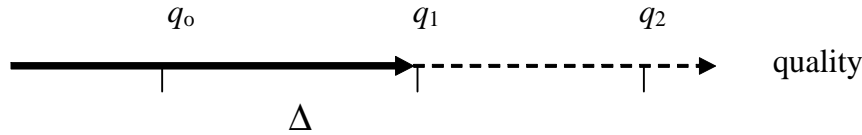


Figure 1: A quality ladder

asymmetry may vanish (depending on the structure of patent law) if every innovator is eventually in both positions, first a licensee and then a licensor, as may occur on a “quality ladder.” A quality ladder is a model used by economists to describe a sequence of products, each of which improves on its predecessor. Each improver is first in the position of having the latest and best product, and is later displaced.

Figure 1 shows a quality ladder where the highest quality so far is q_1 , and the previous quality was q_0 . The conceit is that q_1 could not have been invented without either learning the q_0 technology from that inventor, or replicating the prior knowledge that q_0 represents. If q_1 gets the benefit of learning, then it owes a moral debt to q_0 and to all its predecessors. But instead of being *rewarded* for the contribution that q_0 and its predecessors make to q_1 , the proprietor of q_0 and its predecessors will be *punished*, since q_1 will displace them from the market. It is small solace to the proprietor of q_0 that the proprietor of q_1 will eventually suffer the same fate.

When the two stages of research become an ongoing sequence, the division of profit between leaders and followers might not matter very much, depending on the precise structure of patents. If all improvements infringe prior patents, each innovator will eventually be in each position, first a follower, then a leader. What matters for incentives to invest is the total flow of profit, and whether it is large enough to support the sequence of improvements.¹⁹ However, the matter is a bit more complex when some innovations escape infringement altogether, as we will see.

Breadth and patentability can be organized as the following four types of protection.

1. (Maximalist) All improved products are patentable, and every improvement infringes every prior patent.
2. (Maximalist with a limitation on breadth) All improvements are patentable. Every “small” improvement infringes every prior patent, but “large” improvements do not.
3. (Minimalist) To be patentable, the improvement must be “large.” No improvements are infringing.

¹⁹This is a little overstated. There may be deadweight losses due to licensing and also due to delay in receiving the reward. Licensing that occurs ex post is less valuable due to both discounting and uncertainty. It also carries transactions costs. I view these losses as second order relative to the main point, which is that it is the average flow of profit that mainly determines incentives. Transactions costs are discussed below.

4. (Enhanced minimalist) To be patentable, the improvement must be “large.” Nonpatentable improvements are infringing, but patentable improvements are not.

To describe the four structures, we need a stylization of the patentability standard and the breadth standard. Let Δ^P be a minimum improvement such that the improved product is patentable, and let Δ^B be a minimum improvement such that the improvement does not infringe.²⁰ That is, an improvement of size Δ infringes if $\Delta < \Delta^B$ and is patentable if $\Delta \geq \Delta^P$.

I have labeled the first system “maximalist” because every improvement has two forms of protection. Absent infringement, the improved product is protected by its patent. But absent the patent, the improved product could in any case be protected by exclusive licenses on prior patents that it infringes. Because it takes the agreement of at least two patent holders to bring the product to market, this is a situation of “blocking patents.” In the simple modeling language I have developed, the maximalist structure corresponds to $\Delta^P = 0$, $\Delta^B = \infty$.

The limited maximalist structure modifies the maximalist structure by making the large improvements noninfringing. This has the effect that an improver with a large enough improvement can displace the previous market incumbent, even if the incumbent’s patent has not expired. This is modeled as $\Delta^P = 0$, $0 < \Delta^B < \infty$.

I have labeled the third scheme “minimalist” because, although large improvements are protected by patents, small improvements are not protected at all, either by their own patents or by infringing others. This entails that $0 < \Delta^P < \infty$, $\Delta^B = 0$.

The enhanced minimalist structure modifies the minimalist structure by making the small, unpatentable improvements infringing, so they can be protected by exclusive licenses, even if not by patents. This entails that $0 < \Delta^B \leq \Delta^P < \infty$ or $0 < \Delta^P \leq \Delta^B < \infty$.

There is one additional structure of interest because, although it cannot be replicated by patents, it has appeared in at least two *sui generis* protection regimes. Choosing one of them, I will call this “the SCPA rule.”

5. (The SCPA structure) Infringement and protection are alternatives. Small improvements are infringing and not protected, while large improvements are noninfringing and protected.

The two regimes that this structure approximates are the Semiconductor Chip Protection Act of 1984 (SCPA) and the Plant Variety Protection Act (PVPA). In those regimes, a large enough advance is protected and noninfringing, and any smaller advance is infringing and not protected.²¹ Thus, an improvement is either protected or infringing, but not both.

²⁰This is the notion of breadth called “leading breadth” by O’Donoghue, Scotchmer and Thisse (1998), but here I will simply use the term breadth.

²¹SCPA, which is codified as 17 U.S.C. sections 901-14, provides an explicit exemption for reverse engineering, and

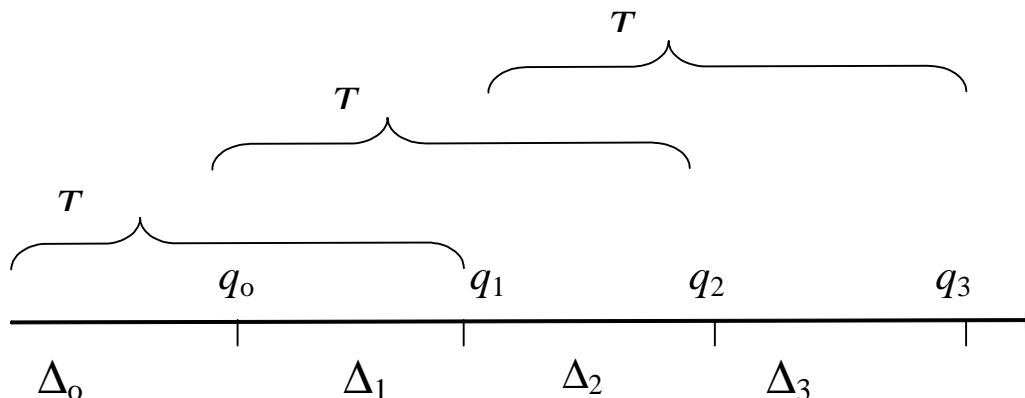


Figure 2: Maximalist: Every improvement is infringing.

The SCPA structure provides different possibilities than patent law because there is only one policy lever, the combined standard that divides infringing and unprotected improvements from noninfringing and protected improvements. As I will discuss below, this is not equivalent to the enhanced minimalist structure, even though it seems it should be when $0 < \Delta^B = \Delta^P < \infty$.

In all four patent structures, and analogously in the SCPA structure, an improved product displaces the previous product if and only if it is *patentable and noninfringing*.²² The expected time until such an innovation occurs is what O’Donoghue, Scotchmer and Thisse (1998) called “effective patent life.”²³ It is determined by the patent’s breadth.

I will argue that the distinctions among the five structures are not so much in whether they privilege leaders or followers, but rather in whether the patent life is determined by the statutory patent life or by the breadth of prior patents, and in whether the scheme privileges small improvements and large improvements differently. I will also argue that these schemes have different merits

allows the reverse engineer to market his product, unless it is “not original” (section 902(b)). The reverse engineer is allowed to “incorporate the results of such conduct in an original mask work which is made to be distributed.” (section 906(a)(2)). The Plant Variety Protection Act, enacted in 1970 and amended in 1994 (see USDA 2005), grants protection for a variety that is new and “distinct” (section 4.2). The right is to exclude others from “selling,... reproducing, ...importing,...exporting.,, or.. producing (as distinguished from developing) a hybrid or different variety.” (section 83(a)(1). The act thus preserves an option for experimentation to develop a new seed, which is itself protected and not infringing if it passes a threshold of distinctness. This structure has some elements in common with copyright law, where the essence (with exceptions) is that an infringing work is not copyrightable. See Lemley (1997). Except where there are compulsory licenses (such as mechanical licenses), there is no implicit “research exemption” that allows a creator to incorporate prior works and negotiate later on the basis of the second creator’s copyright. If the use of the prior work was unauthorized, then if all of the second work is contaminated by the first, the first copyright holder can appropriate the second work. Blocking copyrights would generally only arise if the second work has uncontaminated, copyrighted parts which cannot be appropriated by the first patent holder. (Need citations for all this.)

²²The word “patentable” does not apply to the SCPA structure; one should substitute “protected” instead.

²³In fact, according to patent renewal data, patents rarely last for their entire statutory lives. Even though renewal fees are only a few thousand dollars per year, fewer than half of patents are renewed longer than ten years. See Shankerman (1998).

depending on the nature of the innovative process.

Before proceeding, I pause to say more about the incentive problem we wish to address, and how competition plays out on quality ladders. Quality ladders underlie much of the economic analysis of breadth and patentability standards.

3 Incentives for Improvements on a Quality Ladder

I start by stylizing how the market works once the products have been introduced. For simplicity, suppose that all users have the same willingness to pay for improvements to quality. That is, willingness to pay is the *measure* of quality. Quality improvements may have different magnitudes Δ , where Δ is calibrated as the user's willingness to pay for the quality improvement. I will also assume, with little loss of insight, that the marginal cost of production is zero, i.e., that it costs nothing to make additional copies of the good after the R&D cost of developing the improvement has been sunk. This is a reasonably good approximation for information goods like software, and in any case, is not essential to the argument.

If the current best product and its predecessor are both sold in the market at prices p_o and p_1 , it must hold that $p_1 - p_o = \Delta$ when the quality difference is Δ . That is, the price of the better product is larger by its quality advantage. Further, the price of the inferior product is reduced to marginal cost (zero),²⁴ and all users purchase the higher-quality product.²⁵ Even though the inferior product is not sold, the possibility of selling it constrains the price of the better product.²⁶

Consider the incentive to improve quality. It is not only the price that matters to an improver, but also the length of time the improved product earns profit. A potential improver knows that he himself may soon be displaced. The social bind is the obvious one: Even if there is high social value to improvements, improvers may be scared off by the prospect of rapid turnover.²⁷

²⁴This assumes that the firms compete on price. Economists say that the firms “compete on price” if each firm takes the other's price as given when choosing its own price. We say that they “compete on quantity” if each firm takes the other's supply as given when choosing its own supply. In the situation discussed here, competition on quantity would yield positive prices for both firms, but the difference in price would still be Δ .

²⁵If the price of the inferior product is not zero, the proprietor of the inferior product could reduce price a little and serve the whole market. That is, it must hold that $p_o = 0$ and $p_1 = \Delta$, where p_o is the price of the inferior product at a given point in time, and Δ is the difference in quality between the best product quality q_1 and the previous quality q_0 . Unless the entire market is supplied with the better product, the proprietor of the better product can reduce price a little and remedy that situation.

²⁶Some parts of this story are robust, and some are not. If users have different willingness to pay for quality, then we cannot measure the quality difference by willingness to pay, and the price difference is harder to characterize. Further, if the two products are proprietary, both will be sold in the market, the inferior product being sold at a lower price to the users with lower willingness to pay for quality. See, generally, Gabscewicz and Thisse (1982), and O'Donoghue, Scotchmer and Thisse (1998) for how the argument can be used in the context of intellectual property.

²⁷In general, the profit earned by the improvement will be much smaller than its social value. Because an improvement of size Δ creates a value that lasts forever (every future innovation builds on it), its social value is Δ/r , where r is the interest rate. But the profit is only Δ times the length of time in the market (appropriately discounted, like

In principle, patent design can solve this problem in two ways. First, it can increase the quality gap between competing products. Second, it can extend the expected time until a market incumbent is displaced. The interesting feature of the problem is that there may be a contradiction between increasing the profit and promoting progress. If the average size of improvement is Δ , and there are τ years between improvements, then the rate of progress is Δ/τ per year, while the profit of each improver is $\Delta\tau$. If the average size Δ of improvements increases, both profit and the rate of progress increase. But if the time between improvements increases (τ increases), profit increases but the rate of progress decreases.

Moreover, there is a question of whether it is more efficient to make progress in large steps or small steps. Suppose, for example, that the R&D cost of an improvement Δ is given by $c(\Delta)$. If the ratio $\Delta/c(\Delta)$ is large when Δ is large, then it is efficient to make progress in large increments Δ . The minimalist patent structure would work fairly well, since it takes a large increment of progress to get a patent, and the market will proceed with a sequence of large improvements, each (possibly) displacing its predecessor, but providing considerable profit in the meantime.

The harder problem is when the ratio $\Delta/c(\Delta)$ is large for small values of Δ . Then it is cheaper to make progress in small steps, assuming that the same total progress can be made in more or less the same length of time. If a different firm is likely to achieve each improvement, the small steps could erode profit. Not only is each small step quickly superseded, but the improver earns only a small flow of profit before that happens.

My inquiry is how patent law should be structured to give enough reward to improvers, while making progress efficiently, which may entail that many improvers make small improvements and that the improvements are disclosed, each becoming a basis for future progress. As I have pointed out, there is a legal remedy to the contradiction that disclosure of frequent small improvements may be good for progress but bad for profit. The remedy is infringement. Then the small steps will be disclosed (shared) in a patent, and consolidated through licensing, which will avoid profit erosion. As I will develop more below, licensing allows an improvement to be profitable without ending the profit of the prior innovation. However, the infringement trick can play out in various ways. I will consider the five structures given above.

4 Four Ways to Organize Patentability and Breadth

The maximalist structure: Living out the statutory life

Figure 2 shows the maximalist structure, where patents are very broad, and every improvement infringes every prior patent that has not expired. The effective life of every patent is its statutory life, but the flow of profit at each point in time depends on the number of improvements that are

the social value), which will generally be much less.

consolidated by license, and the distance from the last expired patent. In all the structures except the maximalist, the effective patent life may be terminated by a noninfringing and patentable improved product before the end of the statutory life.

In figure 2, there are two improvements consolidated by license at each point in time. For example, in the period between the invention of q_1 and the invention of q_2 , the owner of q_1 is the incumbent in the market, licensing the patent on q_0 , so the flow of profit is $\Delta_0 + \Delta_1$. In the period between the invention of q_2 and the invention of q_3 , the owner of q_2 is the incumbent in the market, licensing the patent on q_1 , so the flow of profit is $\Delta_1 + \Delta_2$.

This structure seems to embody the intention of patent law, but has the defect that the total profit depends on the rate of making improvements. Suppose the improvements come at a fast clip. Then, instead of including two improvements on average, as in figure 2, each umbrella bracket would include more improvements. The improvements that are eventually consolidated under the control of q_2 might become, for example, $\Delta_0 + \Delta_1 + \Delta_2$, with a correspondingly higher price. If the cost per improvement is more or less the same, regardless of how fast the improvements come, then improvers will be much better rewarded when the improvements arrive at a fast clip. The high reward can lead to inefficiencies, either unnecessary deadweight loss or wasteful duplication of costs in patent races.²⁸

The problem with the maximalist structure is that profit is hard to control. If every improvement infringes every prior patent, there is no policy lever under the control of the patent office or courts that could modify the profitability of a patent according to market circumstances or cost. The other three patent structures might do better, because the effective patent life is endogenous, and the effective patent life depends on policy levers that are under the control of policy bodies such as courts.

In any case, as I have mentioned, patent renewal data show that patents rarely live out their statutory lives.

The maximalist structure with a limitation on breadth

If the maximalist structure is modified by limiting the breadth, patents may terminate early due to a noninfringing improvement. Broader patents have longer effective lives, which are nevertheless capped by the statutory life. The maximalist structure is where breadth is “infinite” (everything infringes), and the effective patent life is the statutory life.

Improvements such that $\Delta > \Delta^B$ are noninfringing, where we now assume that $\Delta^B > 0$. In figure 3, the product q_1 embodies a small improvement Δ_1 , where Δ_1 is smaller than Δ^B , and infringes q_0 . The product q_2 , with the improvement Δ_2 , also infringes q_1 . The product q_3 , with

²⁸See chapter 2 of my book *Innovation and Incentives* for a discussion of why it can be inefficient to give very high rewards to innovators.

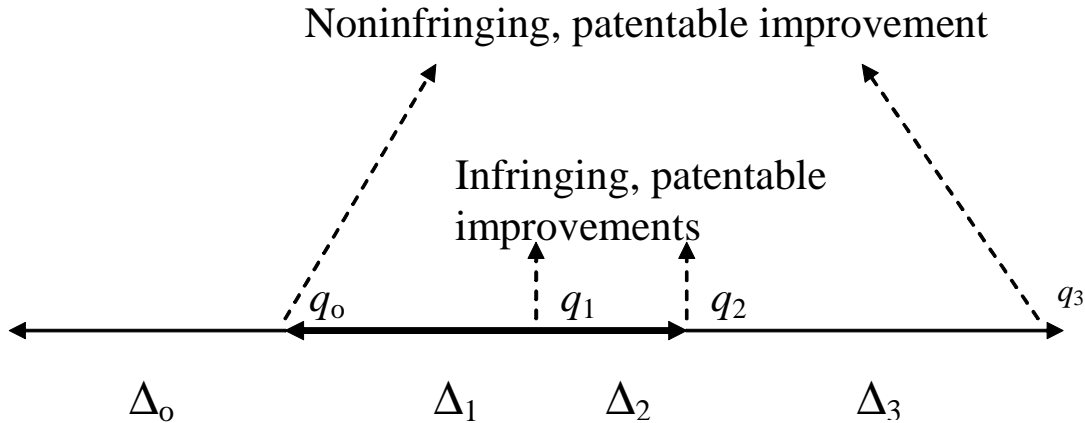


Figure 3: Limited Maximalist: All improvements are patentable; small improvements are infringing.

the larger improvement Δ_3 (larger than Δ^B) does not infringe. Each infringing improvement is licensed. The three products (q_0, q_1, q_2) end up under the control of a single proprietor, assumed for concreteness to be the last noninfringing patentee, with a quality gap (price) equal to $\Delta_0 + \Delta_1 + \Delta_2$. This lasts until the noninfringing product q_3 appears and replaces them in the market, selling at price equal to Δ_3 . Each patentable, infringing improvement raises the bar for avoiding infringement, and extends the effective patent life of the incumbent who is accumulating improvements under license.

Provided the patents are sufficiently narrow, the effective life will be shorter than the statutory life. (If not, the maximalist structure is still in effect.)

The minimalist structure: A patentability standard with no breadth (no infringement)

In the minimalist structure, there is no consolidation of rights by licensing, since improved products never infringe prior patents. The system operates entirely by setting a high bar for patentability, namely, Δ^P . Then, assuming that innovations cannot be protected as trade secrets, no one will invest in unpatentable improvements. If an innovator wants to supplant the incumbent, he must do so with a large improvement, $\Delta \geq \Delta^P$. A smaller improvement, being both unpatentable and noninfringing, could be imitated by the prior innovator or someone else.

The objective of creating a high bar for patentability is to make investors more ambitious than they otherwise would be. Doing so can increase each incumbent's flow of profit by ensuring a larger market gap, hence price, and can lengthen the time between improvements.

But the latter point raises the contradiction identified above. How can a long delay between improvements be good for progress? Further, how can it be good for progress if firms must hoard their smaller steps of progress rather than disclosing and sharing them to avoid duplication? For

example, suppose the ratio $\Delta/c(\Delta)$ is larger for small steps than for large steps, as discussed above. Then, even if each firm makes progress efficiently by creating many small steps (each smaller than the patentability standard), the small increments to knowledge will not be patented and disclosed, hence not shared among the firms. This can be privately optimal but will be socially wasteful if it leads to duplicated efforts.

Perhaps paradoxically, the minimalist structure can also create a danger of entrenched market power. If the incumbent can invest efficiently in small improvements, and use the knowledge as a trade secret, it becomes very hard to displace him from the market. His incumbency may never end. The virtue of rivalry is that firms are eventually displaced, creating competition between proprietors of noninfringing, protected products.

I reiterate, though, that there is a positive side, namely, that a patentability standard can make innovators more ambitious. Firms can become more ambitious in two ways. They can invest more heavily to accelerate their discoveries, or they can target a larger improvement. A high patentability standard can cause the firms to target a higher level of knowledge, which then causes them to invest more resources in order to avoid the slowdown that would otherwise be caused by the more ambitious target (O'Donoghue, 1998).²⁹ Alternatively, reducing the likelihood that a successful innovator can get a patent may cause the firms to invest more in order to compensate (Hunt, 2004).³⁰

The enhanced minimalist structure

In the minimalist structure, no improvements are infringing and only large improvements are patentable. Small improvements are not protected. But if small improvements are infringing, the small improvements can be protected with exclusive licenses. That is the rationale for considering an enhanced minimalist structure where the nonpatentable improvements infringe the prior patent.

The limited maximalist structure and enhanced minimalist structures share the feature that a small improvement is protected because it infringes a prior patent. An improver can be protected under an exclusive license. They also share the feature that the patent life is endogenous, and ends when a noninfringing, patentable improvement is made. There is much similarity. However, I argue that the enhanced minimalist structure has a defect that is avoided by the limited maximalist

²⁹In the model of O'Donoghue (1998), firms choose both the targeted objective and a rate of investment. The high standard will affect both. In that model, firms will always target the efficient level of improvement, but invest too little in achieving it. Although O'Donoghue warns that this result is not robust, the model does show that, by increasing the patentability standard, firms can be induced to invest at a higher rate and achieve the innovations sooner.

³⁰Hunt (2004) also studies a notion of nonobviousness, but in his model, the nonobviousness standard plays a completely differently role. The investing firm faces a "lottery" on the size of the improvement. The size of the improvement is not under his control. A higher nonobviousness standard reduces the probability that the improvement will be patentable, and thus reduces profit. The nonobviousness standard should be higher where it is cheaper to make progress. This is a way of saying that where progress is cheap. The nonobviousness standard is a way to control rewards.

structure. The enhanced minimalist structure can create a “gap” where improvements are not protected at all. That is, smaller improvements might be protected through infringement, and larger improvements might be patentable, but there is a gap in the middle where neither applies.

Clearly this can happen if $0 < \Delta^B < \Delta^P < \infty$. A middle-sized improvement Δ that satisfies $\Delta^B < \Delta < \Delta^P$ is neither infringing nor patentable, while smaller and larger improvements are protected either by infringement (exclusive licenses) or patents. Since there is no obvious rationale for excluding middle-sized improvements from protection, this is probably not the optimal structure.

But oddly enough, a gap can also appear when $0 < \Delta^P \leq \Delta^B < \infty$. Essentially this is because the breadth Δ^B is measured from the prior *patented* knowledge, while the patentability standard Δ^P would logically be measured from the prior *disclosed* knowledge. Knowledge can be disclosed in licensed, infringing improvements that are not patented. This is illustrated in figure 4.

Figures 3 and 4 describe the limited maximalist and enhanced minimalist structures, respectively. The figures are based on the same sequence of innovations, and the same breadth standard Δ^B . In both figures, q_0 is a noninfringing and patentable product, and q_1 is the next improvement. The improved product q_1 is infringing in both figures, that is, $\Delta_1 < \Delta^B$, and is therefore licensed to the proprietor of q_0 . The proprietor’s price in the market is then $\Delta_0 + \Delta_1$.

Now suppose that q_2 appears, and that q_2 also satisfies $\Delta_2 < \Delta^B$. In figure 3 (the enhanced maximalist structure), q_2 infringes the patent on q_1 . Therefore q_2 is also licensed to the proprietor of q_0 (who holds the rights on q_1), and the proprietor’s profit increases by an amount equal to the increment to quality Δ_2 . This is how q_2 is protected.

But in figure 4 (the enhanced minimalist structure), q_2 does not infringe q_1 , since q_1 does not have a patent. As drawn, q_2 also does not infringe q_0 because the quality gap between q_2 and q_0 is larger than Δ^B . Further, q_2 is not patentable if $\Delta_2 < \Delta^P$, and is then not protected at all. Any firm could imitate q_2 , thus competing the market price to zero. No one would invest in q_2 unless it could be protected as a trade secret.

The problem in figure 4 is that the licensed improvement q_1 raises the bar for patentability, since q_1 is “prior art”, but q_1 does not raise the bar for infringement, since there is no patent on q_1 to infringe. This is what creates the gap.

Thus, although the limited maximalist structure and the enhanced minimalist structure have much in common, the limited maximalist structure avoids the problem that an improvement may be stranded with no protection at all. In this sense it dominates the enhanced minimalist structure.

The SCPA structure

The limited maximalist structure also has much in common with how rights are structured under two *sui generis* protection regimes, the Semiconductor Chip Protection Act of 1984 (SCPA),

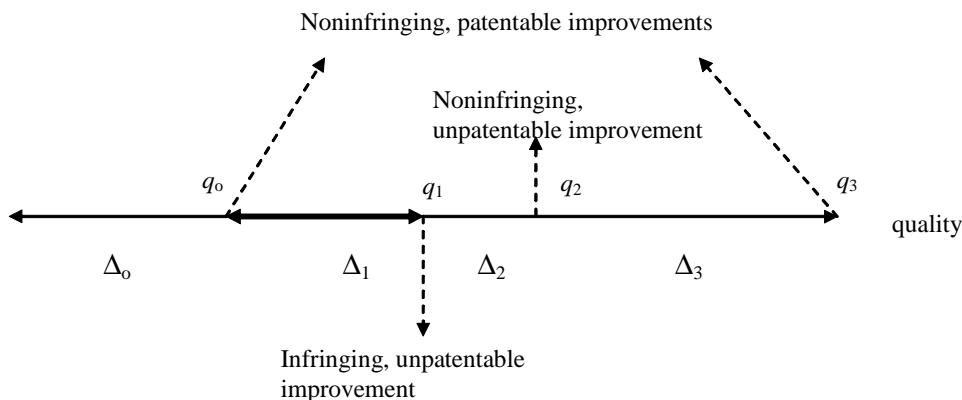


Figure 4: Enhanced minimalist: There may be a gap in protection.

and the Plant Variety Protection Act (PVPA). However, those regimes are much simpler because there is a single policy lever rather than two. In those regimes, an advance is either protected or infringing, but not both. The single policy lever is the minimum advance such that the improvement becomes noninfringing and protected rather than infringing and unprotected. Small increments to progress are infringing but not protected, while large increments are protected but not infringing.³¹

The SCPA structure cannot be replicated by the patents, except where there is no licensing. With no licensing, the SCPA structure would be replicated by setting $\Delta^B = \Delta^P$. But figure 4 can describe the case $\Delta^B = \Delta^P$. Figure 4 shows that, as soon as an infringing improvement is licensed, the bar for patentability becomes higher, whereas the bar to escape infringement does not. This creates a gap of unprotected qualities, which never occurs in the SCPA structure. In the SCPA structure, when an infringing improvement is licensed, there is a commensurate increase in the bar for both escaping infringement and claiming protection.

³¹SCPA, which is codified as 17 U.S.C. sections 901-14, provides an explicit exemption for reverse engineering, and allows the reverse engineer to market his product, unless it is “not original” (section 902(b)). The reverse engineer is allowed to “incorporate the results of such conduct in an original mask work which is made to be distributed.” (section 906(a)(2)). The Plant Variety Protection Act, enacted in 1970 and amended in 1994 (see USDA 2005), grants protection for a variety that is new and “distinct” (section 4.2). The right is to exclude others from “selling,..., reproducing, ...importing,...exporting.,, or.. producing (as distinguished from developing) a hybrid or different variety..” (section 83(a)(1). The act thus preserves an option for experimentation to develop a new seed, which is itself protected and not infringing if it passes a threshold of distinctness. This structure has some elements in common with copyright law, where the essence (with exceptions) is that an infringing work is not copyrightable. See Lemley (1997). Except where there are compulsory licenses (such as mechanical licenses), there is no implicit “research exemption” that allows a creator to incorporate prior works and negotiate later on the basis of the second creator’s copyright. If the use of the prior work was unauthorized, then if all of the second work is contaminated by the first, the first copyright holder can appropriate the second work. Blocking copyrights would generally only arise if the second work has uncontaminated, copyrighted parts which cannot be appropriated by the first patent holder. (Need citations for all this.)

5 Transactions costs

Before turning to which structure is optimal, or in what circumstances, I comment on how the feasible options change if there are transactions costs in licensing. If the transactions costs are high enough, only the minimalist structure survives.

Transactions can be in jeopardy for many reasons. Most obviously, a party might think it can do better by not making the deal. However, that is not generally a failure due transactions costs, but rather a recognition that the deal must respect bargaining positions. Outside options should affect the terms of the deal, but not whether the deal takes place.

By transactions costs, I mean that real resources are required to make a deal. Then deals that would otherwise improve the profit of both parties relative to their bargaining positions might be stymied. For example, the costliness of enforcing nondisclosure agreements would be a transaction cost.

If transactions costs are so severe that licenses will not occur either *ex ante* or *ex post*, the only feasible structure of rights is what I have called the minimalist structure. “Large” improvements are patentable and noninfringing, and small improvements, being neither infringing nor patentable, will not be brought to market.

A less severe form of transactions costs is that deals can be made *ex post* but not *ex ante*, for example, because nondisclosure agreements are ineffective. Then an improvement that is infringing but not patentable might be rejected, because the improver would be in a weak bargaining position *ex post*. After sinking costs, he could not bring the improvement to market without a license, and could not prevent the previous patent holder from adopting his improvement without paying. The previous patent holder might appropriate all the value, so he cannot even recover his sunk costs. Patentability improves the *ex post* bargaining position of improvers, and may lead to improvements that would otherwise be stymied, although the improver is still in a weaker bargaining position than if he were negotiating *ex ante*, since he has already sunk costs.

This is a way of saying that transactions costs impinge on the innovative process. The minimalist structure will not support an innovation strategy where firms make progress in small steps that are disclosed to rivals. A firm might choose to work in small steps, but will hoard its knowledge until the number of steps accumulates to a patentable innovation, if not longer. Progress is then slower than if the steps are shared, because rivals can end up duplicating knowledge that already exists.³²

Transactions costs may change our prescription for patent law because they change the struc-

³²This is the basis of the Scotchmer and Green (1990) paper on disclosure. If it takes two steps of progress to receive a patent, then a firm with one step of progress will not patent and share it. Rivals may therefore be hard at work duplicating progress that has already been made, but not shared. A good incentive system would encourage disclosure at intermediate stages, to avoid this waste.

tures that are in the first instance feasible. Namely, in the extreme form, only the minimalist structure remains.

6 Trade Secrets and Disclosure

Another important institutional reality is that innovators can keep their innovations secret instead of patenting them. An innovator will choose secrecy only if secrecy protects longer or more effectively than a patent. The option to use trade secrecy increases the length of protection and increases the profitability of the innovation; otherwise the innovator would choose to patent.³³

The fact that trade secrets increase protection can obstruct the well thought out objectives of patent law. Patent law is (or should be) carefully designed to balance the reward to an innovator against the benefits to end users and to follow-on innovators. Trade secrecy is not “designed” at all, except for special rules, such as those that govern reverse engineering. If trade secrecy gives protection that is too strong by the standards that would be used to judge patent law, there is not much to be done. One could strengthen patents to encourage patenting rather than secrecy, but that only formalizes the too-strong protection provided by secrecy. It may nevertheless be the best of bad options, since it provides disclosure benefits that secrecy does not.

The latter is the important defect of trade secrecy that I wish to emphasize. In the cumulative context, the problem is not so much that trade secrecy protects longer than a patent, but that it protects without disclosure.

I pause here to comment on the role of disclosure in patent law. Disclosure preserves and transmits knowledge, and gives a basis for the claims in the patent. At the same time, the patent holder has an exclusive right to use the disclosed knowledge. The story told above – that rivals can build the next step of progress from the patented base of knowledge – implicitly asserts some kind of research exemption. A rival can use the knowledge to develop an improvement, even if he needs a license to commercialize it. This is probably a good description of how patents operate in practice, which is why I have adopted it, but it is not how patents operate in theory. There is slippage between theory and practice.³⁴

If there were no such slippage, patenting and trade secrecy would be more similar. Trade secrecy is also imperfect, but it starts from the opposite assumption, namely, that the knowledge is not disclosed and rivals cannot use it, either for commercialization or for further progress, as long as the

³³There is convincing survey evidence that firms depend heavily on trade secrecy (Cohen, et al, 2000), but that does not answer the question whether trade secrecy subverts the well designed intentions of patent law, or instead compensates for intellectual property rights that are too weak.

³⁴These conundrums have been well articulated by Rebecca Eisenberg (1989) in a paper that began as important and evolved to essential reading, as technologies have become more complex and cumulative.

secret is legitimately kept.³⁵ It is the secrecy that prevents rivals from building on the knowledge.

Nondisclosure agreements are a vehicle that tries to bridge the gap between trade secrets and the automatic disclosure of patents. If a nondisclosure agreement is signed, the owner of the secret can, in principle, negotiate without fear that the secret will be stolen, just as a patent holder can. However, this does not go the whole distance. Even if nondisclosure agreements can protect against theft, they do not serve the same purpose as disclosure in a patent. The disclosure in a patent is not tied to a specific contract with a specific firm, and can therefore stimulate imagination across a wide set of potential innovators. Without first imagining an improvement, there is no reason to negotiate a nondisclosure agreement.³⁶ Nondisclosure agreements have a chicken-and-egg problem. As I will argue in the next section, the importance of stimulating a flow of ideas depends on the nature of the protected knowledge, and the nature of the innovative process.

Secrecy is obviously more effective, and consequently more erosive of patenting, when the trade secret can be used (marketed) without being disclosed in a way that is useful to rivals. A secret will remain in effect, with its owner collecting proprietary profit, until the product is duplicated or replaced in the market, possibly because the secret is breached, or possibly because a rival discovers how to replace the product, perhaps with a superior one.³⁷ This is likely to be a time-consuming process. The cost of supplanting an incumbent will be high because knowledge is not fully shared, but the reward to success is also high, because the improver's own time in the market is likely to be long. Nevertheless, progress is inefficient due to the lack of disclosure. A rival displaces an incumbent by duplicating his knowledge before building on it.

Because trade secrecy does not lead to licensing of intermediate progress, the closest patent structure is again minimalist, where the patent holder remains in the market until replaced, but without accumulating steps of progress provided by other improvers, and without sharing his own progress with rivals through patenting. But the trade secrecy regime is less efficient than the minimalist patent structure because rivals cannot work from the knowledge base provided by the previous innovator. Instead, a rival must duplicate the progress that has been made before, but kept secret.

This quasi-minimalist trade secrecy regime is an echo of what I argued for transactions costs, namely, that with enough impediments to licensing, the only possibility is equivalent to the minimalist patent structure. As with transactions costs, I am assuming that secret intermediate progress cannot be shared under license.

³⁵The common law of trade secrecy imposes obligations on employees to keep the secret, as long as the owner has made efforts in that direction. See But if the trade secret is independently discovered rather than being breached, it cannot be recovered.

³⁶Maurer and Scotchmer (2007) make this argument in defense of the open source movement. Computer software has the odd feature of being protected by copyrights and patents, but without disclosure in either regime. Imagination may therefore be stifled. Like patents in other contexts, open source protocols create disclosure together with protection.

³⁷Need an authority saying what trade secrecy protects against.

7 What can we learn about optimal patentability and breadth standards?

So far, I have said little about social welfare, or about the tradeoff between deadweight loss and innovation. A complete analysis should consider not only how patentability and breadth standards should be structured, but also the size of rewards.

Knowledge, once discovered, lasts forever. In the cumulative model discussed above, each improvement lays a foundation for all future improvements. If an improvement of size Δ were removed from the sequence of improvements, then at each time in the future, the highest quality would be smaller by Δ . The social value of an improvement therefore lasts forever. Discounted at rate r , and if there is no reduction of value in raising the funds to pay for it, the social value of the innovation is Δ/r .

Should we then give a reward that is equal to Δ/r ? At least two inefficiencies could arise. Depending on the nature of the innovative environment, and assuming that R&D costs are considerably smaller than the social value, this could lead to a wasteful patent race. Even if not, there is a deadweight loss due to proprietary pricing, which more or less scales with the size of the reward. By “deadweight loss” I refer to the exclusions on use that arise because some users who would be willing to pay competitive prices are not willing to pay proprietary prices.³⁸ The per-period social value is smaller during the period of protection than afterwards, because of the deadweight loss that accompanies proprietary pricing.

My focus in this paper is not mainly on the optimal size of rewards, although that is also at issue, but rather on how rewards should be structured. The reward structure must address the contradiction between rapid turnover as an engine of progress, and the profit erosion that results from it, and must have enough levers so that different industries can be treated differently.

There are clear defects to both the minimalist and maximalist structures. In the maximalist structure, each patent lives out its statutory life. As a consequence, policy bodies such as courts or the patent office have very little flexibility to adjust profit to the needs of the industry. A high rate of improvement will lead to a high flow of profit, due to the consolidation of many licensed improvements before the patent expires. Innovators may therefore be overrewarded, leading to unnecessary deadweight loss or wasteful patent races. The statutory life can be changed by Congress, but innovators will still be under-rewarded or over-rewarded in environments with different rates of improvement.³⁹

³⁸The simple model above does not model deadweight loss explicitly. In the similar model of O’Donoghue, Scotchmer and Thisse (1998), deadweight loss takes the form that users with low willingness to pay for quality always end up consuming a lower-quality product, even though a higher-quality product is available. O’Donoghue (1998) models deadweight loss by asserting it, without being specific about its source.

³⁹A partial solution is to create different protection regimes for different subject matters, as was done in the Semiconductor Chip Protection Act of 1984, which provided for 10 years of protection.

In contrast, if the effective patent life is controlled by breadth, as in the limited maximalist system, the effective patent life is largely under the control of patent examiners and courts. The law does not seem to have taken advantage of this potential.⁴⁰ The ability to fine-tune effective patent life according to the realities of a particular industry is an important advantage over the maximalist structure, and suggests that the maximalist structure with limited breadth is superior.

Another important feature of the limited maximalist system is that large improvements have a profit advantage. They do not pay licensing fees to earlier innovators, but share in the profit of later, infringing improvements. For this reason alone, regardless of the price advantage that comes from a larger market gap, it is more advantageous to make a large, noninfringing improvement than to make a small, infringing improvement, even if both are patentable. The limitation on breadth skews the inventive effort and the total profit flow toward large, noninfringing improvements. Whether this is efficient depends on whether it is efficient to make progress in small, shared steps, and whether it is hard to entice improvers to make them.⁴¹

In the minimalist structure, a firm may invest in small, infringing steps if that is efficient, but has an incentive to hoard them until they accumulate to a patentable improvement. The hoarding imposes a social cost, since rivals will either drop out or duplicate progress that has not been disclosed.⁴² In the absence of transactions costs, I see no reason to choose this structure.

Where transactions costs undermine licensing, I have argued that the minimalist structure may arise in any case, whether or not patent law is structured that way. Suppose, for example, that the formal structure is the limited maximalist structure, where every improvement is patentable, but small improvements infringe a prior patent. No one will invest in the small, infringing improvements if they cannot be licensed. Instead, investment will be directed to improvements that are large enough to avoid infringement, just as if the patent law had been structured that way to start with. The breadth standard in the limited maximalist structure then plays the role of the patentability standard in the minimalist structure.

⁴⁰Note to self: Is this true? I need a citation discussing whether courts have taken advantage of this. Does *Festo* have any elements of this? See Merges' case book? It is, as lawyers say, hard to "prove a negative." But I seem to need an authority for this.

⁴¹In the model of O'Donoghue, Scotchmer and Thisse (1998), where opportunities to make improvements occur as a random process, and the size of the improvement is not under the control of the improver, it is the willingness to invest in small improvements that determines the rate of progress. Thus, the skewing of profit toward large improvements involves social waste. Given the rate of progress, the required profit could be less if more equally shared. They argue that the profit will be more equally shared between large and small improvements in the maximalist structure than in the limited maximalist structure.

⁴²This is shown by Scotchmer and Green (1990), who study a two-step innovation process, and compare regimes where it takes one or two steps to be patentable. They show that when it takes two steps, the firms may nevertheless hoard the first step of knowledge in order to keep an advantage in achieving the big payoff at the end. In their case, the hoarding takes the form of not marketing the intermediate step, which they assume would disclose it, but the same logic applies to licensing. Hunt (2004) and O'Donoghue (1998) study the minimalist structure, but simply assume that firms will try to achieve the minimal patentable improvement, rather than making progress in small steps if that is efficient.

Without considering trade secrecy, it thus seems to me that there are strong reasons to choose the maximalist structure, possibly with limitations on breadth. Where transactions costs are not high, the limited maximalist structure lets small steps of progress come to market, earning modest profit under license, without threatening the profit of the prior innovator. Breadth should be lower where the rate of idea formation is low. Otherwise, it will take a long time before there is an idea that jumps the breadth hurdle, and improvers may be overrewarded.

Trade secrecy has the pernicious effect of closing off disclosures that would otherwise allow rivals to build on knowledge created by the previous improver. Perhaps counterintuitively, the incentive to keep an innovation secret is even stronger if the proprietor can negotiate licenses under nondisclosure agreements. If so, many increments of progress can be consolidated under the proprietor's control. Each undisclosed improvement makes it harder to unseat the incumbent.

There is not much that patent law can do about this, except to encourage disclosure by providing stronger protection, for example, more breadth and a higher patentability standard. In both the patent and secrecy regimes, the proprietor can consolidate improvements under license, each one creating a higher hurdle for rivals. In both cases, it is reasonable to think that the proprietor will eventually be unseated. But to unseat the proprietor of a trade secret, the rival must leap a higher hurdle. He must first duplicate the knowledge that was secret. Thus, it is easy to see how trade secrecy could last a long time.

Beyond this, the best way to structure rights depends on the nature of the innovative process. Economists have put forward several models.⁴³ The one that mostly underlies my own work⁴⁴, and underlies much of the discussion above, is that ideas for improvements (investment opportunities) occur as a random process. An idea specifies the size of improvement and also the required R&D cost. A potential innovator must have an idea before he has an opportunity to invest.⁴⁵

If good ideas mostly do not crowd out others – if investment in one idea sets the stage to let other innovators have even better ideas – firms should invest in even modest ideas, provided the R&D cost c is less than the social value Δ/r . This is what the breadth standard can accomplish. Modest ideas can be disclosed and used under license without threatening the profit of prior innovators. It is the breadth standard (infringement) that allows this, and the breadth standard that determines the average amount of progress that is consolidated through licensing before a nonin-

⁴³See chapter 2 of S. Scotchmer, 2004 *Innovation and Incentives*, Cambridge, MA (MIT Press).

⁴⁴Especially relevant is O'Donoghue, Scotchmer and Thisse (1998), where ideas are for progress on a quality ladder. But see also Scotchmer (1999), where each idea opens a new market, and Erkal and Scotchmer (2007), where ideas are substitutes in the same market.

⁴⁵If ideas are scarce, an effective innovative environment must embrace all the potential innovators who might have ideas for improvements. This is an important reason for patent disclosures that is not present in the production-function model, where opportunities for investment are common knowledge. There is indirect evidence that something like the ideas model is operative, or at least that opportunities to invest are scarce. If ideas are common knowledge, profit should be dissipated in patent races. That does not always occur, as is evident from the “dot boom” in Silicon value, and even more systematic evidence on the return to R&D spending. See the work cited in chapter 9 of my book *Innovation and Incentives*.

fringing improvement appears. In this innovative environment, it is hard to see a reason to limit patentability by creating a patentability standard. For example, a patentability standard cannot be used to make innovators more ambitious in the improvements they seek, because that is not under their control.

The “ideas” model is not the only way that technology proceeds. Much of the economics literature posits a “production function for knowledge.” This model may also have merit, especially in established research arenas like drug research. The production function expresses how R&D inputs can be turned into R&D outputs. The input is generally labor or money, and the output is generally the time it takes to achieve the improvement. The size of improvement is stochastic in the production function of Hunt (2004), and targeted in the production function of O’Donoghue (1998). It is usually assumed that the production function is commonly known to all potential rivals. Since the production function is common knowledge, there will be patent races, and profit will be dissipated.

The patentability standard plays a greater role in the production-function model. In fact, the production-function models of O’Donoghue and Hunt do not use patent breadth in any important way at all.⁴⁶ The basic idea is that a more stringent patentability standard lengthens time in the market. Because rivals must achieve a larger improvement, they find it more costly and time consuming to unseat the incumbent. This point of view has its own internal logic, but it does not grapple with the fact that progress might be cheaper and faster if rivals were induced to make smaller steps of progress, and to share them in a licensing regime.

The merits of the patent structures discussed in this paper rely on competition between sequential improvers as a way of giving appropriate rewards to improvers while also limiting their market power. This seems like a delicate balancing act. It assumes that incumbents will eventually be unseated by rivals, and that the expected time until the incumbent is unseated can be governed by patentability and breadth standards.

However, a question that has been with us since Schumpeter’s famous (1942) book is whether incumbents are likely to be their own successors. Schumpeter thought monopolists are the engine of progress, possibly because they have more incentive, but surely because they have more money. Economists have been arguing about the relative incentives of incumbent monopolists and rivals ever since.⁴⁷

In the quality ladder model discussed above, for the most part, rivals and incumbents have the

⁴⁶In O’Donoghue, Scotchmer and Thisse (1998), O’Donoghue (1998) and Hunt (2004), it is assumed, as is natural, that the patent reads on every quality in the quality gap created by the improver. However, as stressed by O’Donoghue, Scotchmer and Thisse (1998), the more controversial tool is what they call “leading breadth,” simply called “breadth” in this paper, namely an exclusive right over qualities *larger* than the invented quality. Entry with small improvements is deterred in the models of O’Donoghue (1998) and Hunt (2004) by ensuring that small improvements are not patentable.

⁴⁷Arrow (1962) gave the first argument against Schumpeter, pointing out that incumbents already have a privileged position in the market, and therefore have less to gain by achieving the next innovation.

same incentive to take any given investment opportunity. Regardless of whether an improvement is achieved by the incumbent or achieved by a rival, and regardless of whether the rival licenses to the incumbent or competes with the incumbent, the profit flow generated by the improvement is equal to the quality gap it creates. It is this symmetry of incentives that allows rivals to unseat incumbents, and promotes competition within the confines of the patent structure. Because at each stage there is one incumbent and many potential rivals, incumbents will only rarely be their own successors.

Again, though, an important exception to the symmetry between incumbents and rivals is where improvements are protected as trade secrets. With trade secrecy, a rival can only enter the market by duplicating the incumbent's knowledge. Secrecy may keep the incumbent entrenched for a much longer time than contemplated in the design of patents, possibly because rivals have a cost disadvantage, and possibly because they drop out.

In summary, the economics literature insists that breadth be joined to the debate about patent standards and nonobviousness. The conclusion from that literature that I regard as most robust is that infringement and licensing play a positive role in dynamic innovation environments. Patent breadth is the policy lever that mediates the contradiction between fast turnover and preserving profit. A patentability standard alone is less conducive to sharing knowledge than a patent structure that also relies on infringement and licensing.

On the hand, the sharing of knowledge is by definition limited if firms rely on trade secrecy. Not much can be done about this in the design of patents, except to increase the attractiveness of patenting by increasing the breadth, hopefully with the effect of lengthening market incumbency. The better policy lever is to reduce the efficacy of trade secrecy, for example, by creating a right for reverse engineering.

References

- [1] Australian Patent Office. 2007. *Manual of Practice and Procedures*. www.ipaustralia.gov.au/pdfs/patentsmanual/WebHelp/Patent_Examiners_Manual.htm
- [2] Arrow, K. 1962. "Economic Welfare and the Allocation of Resources for Invention." In R. Nelson, ed., *The Rate and Direction of Economic Activities: Economic and Social Factors*, pp 609-262. National Bureau of Economic Research Conference Series. Princeton, NJ: Princeton University Press.
- [3] Barton, J. 2003. "Nonobviousness," *IDEA: The Journal of Law and Technology* 43, 475-508.
- [4] Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2000. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)." Working paper 7552. Cambridge, MA: National Bureau of Economic Research.
- [5] Denicolo, V. 2000. "Two Stage Patent Races and Patent Policy," *RAND Journal of Economics* 31, 488-401.
- [6] Duffy, J. 2007. "Racing, Timing, and the Patenting Standard," George Washington University, mimeo.
- [7] Eisenberg, R. 1988. "Patents and the Progress of Science: Exclusive Rights and Experimental Use." *University of Chicago Law Review* 56:1017-1055.
- [8] Erkal, N. and S. Scotchmer. "Scarcity of Ideas and Options to Invest in R&D." 2007. Working Paper 07-348. Dept of Economics, U.C., Berkeley.
- [9] European Patent Office. 2007. *Guidelines for Examination*. www.european-patent-office.org/legal/guix/e/c_iv_9_10_4.htm
- [10] *Graham V. John Deere Co.*, 383 U.S. 1 (1966).
- [11] *Graver Tank & Mfg Co v Linde Air Prods Co.*, 339 U.S. 605 (1950).
- [12] Green, J. and S. Scotchmer. 1995. "On the Division of Profit in Sequential Innovation." *The RAND Journal of Economics*. 26:20-33.
- [13] Hunt, R. 2004. "Patentability, Industry Structure, and Innovation," *Journal of Industrial Economics* 52, 401-25.
- [14] Lemley, M. 1997. "The Economics of Improvement in Intellectual Property Law." *Texas Law Review* 75:989-1083.
- [15] *KSR Int'l Co. v. Teleflex, Inc.*, 550 U.S., 127 S. Ct. 1727 (2007)
- [16] Machlup and E. Penrose. 1950. "The Patent Controversy in the Nineteenth Century." *Journal of Economic History* 10:1-29.

- [17] Merges, R. 1993. "Uncertainty and the Standard of Patentability," *Berkeley Technology Law Journal* 7:1-62.
- [18] Merges, R. and R. Nelson. 1990. "On the Complex Economics of Patent Scope." *Columbia Law Review* 90:839-916.
- [19] O'Donoghue, T., Scotchmer, S. and Thisse, J.-F. 1998. "Patent Breadth, Patent Life and the Pace of Technological Progress," *Journal of Economics and Management Strategy*, 7, 1-32.
- [20] Schumpeter, J.A. 1942. *Capitalism, Socialism and Democracy*. New York: Harper and Roe.
- [21] Scotchmer, S. 1991. "Standing on the Shoulders of Giants: Cumulative Innovation and the Patent Law." *Journal of Economic Perspectives* 5:29-41.
- [22] Scotchmer, S. 2004. *Innovation and Incentives*. Cambridge, MA: MIT Press.
- [23] Scotchmer, S. 1996. "Protecting Early Innovators: Should Second-Generation Products be Patentable?" *Rand Journal of Economics*, 27(2), 322-331.
- [24] Scotchmer, S. and Green, J. 1990. "Novelty and Disclosure in Patent Law," *RAND Journal of Economics* 27, 322-331.
- [25] Snyder, D. L. and M. I. Miller. 1991. *Random Point Processes in Time and Space*. New York: Springer-Verlag.
- [26] Strandh, O. 1979. *A History of the Machine*. A&W Publishers, New York.
- [27] U.S. Federal Register 57526. Oct. 10, 2007. "Examination Guidelines for Determining Obviousness Under 35 .S.C. 103 in View of the Supreme Court Decision in *KSR-International Co. v. Teleflex Inc.*
- [28] U.S. Department of Agriculture, Agricultural Market Service. 2005. "Plant Variety Protection Act and Regulations and Rules of Practice." www.ams.usda.gov/science/PVPO/PVPO_Act/PVP2005.pdf.
- [29] U.S. Patent and Trademark Office. 2007. *Manual of Patent Examining Procedure*. www.uspto.gov/web/offices/pac/mpep/documents/0700_716_04.htm