FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall
Philadelphia, Pennsylvania 19106-1574
(215) 574-6428, www.phil.frb.org

Working Papers

Research Department

WORKING PAPER NO. 99-3

NONOBVIOUSNESS AND THE INCENTIVE TO INNOVATE: AN ECONOMIC ANALYSIS OF INTELLECTUAL PROPERTY REFORM

Robert M. Hunt
Federal Reserve Bank of Philadelphia

March 1999

WORKING PAPER No. 99-3

NONOBVIOUSNESS AND THE INCENTIVE TO INNOVATE: AN ECONOMIC ANALYSIS OF INTELLECTUAL PROPERTY REFORM

Robert M. Hunt*
Federal Reserve Bank of Philadelphia

March 1999

* Department of Research and Statistics, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106. Email: bob.hunt@phil.frb.org

The views expressed here are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. The author wishes to thank Chris Chalmers, Stephen Coate, George Mailath, Alberto Trejos, Juuso Valimaki, Rohit Verma, Nikolaos Vettas, Dennis Yao, and participants at workshops at the University of Pennsylvania, the University of Georgia, the Federal Reserve Bank of Philadelphia. Any remaining errors are, of course, my own.

Abstract**:**

U.S. patent law protects only inventions that are nontrivial advances of the prior art. The legal requirement is called *nonobviousness*. During the 1980s, the courts relaxed the nonobviousness requirement for all inventions, and a new form of intellectual property, with a weaker nonobviousness requirement, was created for semiconductor designs. Supporters of these changes argue that a less stringent nonobviousness requirement encourages private research and development (R&D) by increasing the probability that the resulting discoveries will be protected from imitation. This paper demonstrates that relaxing the standard of nonobviousness creates a tradeoff -- raising the probability of obtaining a patent, but decreasing its value. We show that weaker nonobviousness requirements can lead to *less* R&D activity, and this is more likely to occur in industries that rapidly innovate.

1. Introduction

American patent law stipulates that patents may be granted only to inventions that are *useful, novel, and nonobvious*.[1] Taken together, these requirements define a standard of patentability that separates discoveries into those eligible for patent protection and those that are not. The most important requirement is nonobviousness, which one author calls "the ultimate condition of patentability."[2]

A major concern during the 1980s was that discoveries in many high technology industries were often ineligible for patent protection. For example, during hearings on the semiconductor chip industry, one expert testified:

> "...patents generally do not protect the particular topographical layouts created by chip designers. The level of creativity involved in such layout designs does not usually rise to the level required by the patent laws. Most chip layouts fall into the same unpatentable categories as dress designs -- variations on a single idea. Thus the design that makes one's chip layout better than another is generally not patentable."[3]

Congress responded to this problem by creating a new form of intellectual property, called "mask rights," with a weak nonobviousness requirement, to protect chip designs. But other changes in American intellectual property law affected inventors in all industries. For example, a

---

[1] 5 U.S.C. (1988), sections 101-3. Novelty under the patent law stipulates that the person filing the patent is the first person to have made the invention. Nonobviousness is the requirement that the invention be a nontrivial extension of what is already known.

[2] Hence the title of J. Witherspoon's *Nonobviousness-The Ultimate Standard of Patentability*, 1980. See also Merges (92).

[3] Risberg (90), p. 251-2.

reorganization of the federal courts and a series of judicial decisions significantly relaxed the test of nonobviousness used in patent litigation. This paper examines how those changes affect the incentive of firms to engage in R&D.

The value of a patent is affected by its *length*. While the statutory life of a patent is 20 years, we know from empirical research that the economic benefit of a patent is often exhausted before it expires.[4] We call the *economic* life of a patent the amount of time the patented invention generates positive profits. In an environment where unprotected inventions are easily duplicated, the economic life of a patent depends on (1) the rate of innovation, which determines how rapidly potentially competing technologies develop; and (2) the nonobviousness requirement, which determines what proportion of future discoveries will be protected by patents.

A weak nonobviousness requirement implies that most future discoveries will be protected. Today's patent holder will be able to copy the few technologies not protected, but must compete against the rest. Such competition reduces, and eventually eliminates, her profits. Under a strong nonobviousness requirement, only a small proportion of future discoveries is protected. Today's patent holder can copy most of the emerging discoveries. Competing proprietary technologies take longer to accumulate so the patent holder's profits are larger and last longer. Thus, holding the rate of innovation constant, the economic life of patents is increasing in the standard of nonobviousness.[5] This suggests that the value of patents is increasing in the strictness of the nonobviousness requirement.

---

4 See, for example, Mansfield, Schwartz, and Wagner (81).

5 Of course, a change in the value of patents *will* alter R&D activity, and the actual computation of the economic life of patents must take this into account.

The incentive theory of patents argues that R&D activity is positively related to the expected return to R&D. The return to R&D is increasing in the value of patents but decreasing in the probability that discoveries will not qualify for patent protection. Increasing the standard of nonobviousness increases the expected value of patents, but it decreases the probability that a given invention will be protected. We call the increase in the value of patents the *dynamic* effect, and the reduction in the probability of obtaining protection the *static* effect, of raising the standard of nonobviousness. R&D activity increases or decreases, depending on which of these competing effects is stronger. The *conventional wisdom* during the 1980s was that the static effect would dominate any dynamic effects.[6] Under this assumption, relaxing nonobviousness requirements would indeed encourage R&D activity.

We evaluate this assumption using a model of sequential innovation where property rights closely follow the "mask rights" created for the semiconductor manufacturing industry. We demonstrate the existence of a unique standard of nonobviousness that maximizes R&D activity in an industry. Using this critical standard, we can divide the range of possible requirements into regions where either the static or dynamic effect dominates. The share of all possible nonobviousness requirements contained in the static region is a measure of the reliability of the conventional wisdom.

Analysis of the model shows that the static region is smaller in industries that innovate rapidly. In fact, we show that the share of possible standards accounted for by the dynamic region rises when we vary any of a number of parameters that increase the equilibrium rate of innovation

---

[6] In fact, there appeared to be little recognition of any effect on the value of patents. An exception is found in the hearings on the Semiconductor Chip Protection Act. See Section 2.

in an industry. These results suggest that the conventional wisdom is *least* appropriate when applied to rapidly innovating industries.

The remainder of the paper is organized as follows. Section 2 describes the Semiconductor Chip Protection Act and the Federal Courts Improvement Act. Section 3 reviews the literature on patent attributes and the determination of firms' R&D spending. Section 4 introduces the model. Section 5 constructs the symmetric, stationary equilibrium and describes its properties. Section 6 shows how a change in the nonobviousness requirement affects R&D activity. Section 7 concludes.

2. Examples of Intellectual Property Reform

A. The Semiconductor Chip Protection Act of 1984

The semiconductor manufacturing industry is one of the most innovative and research-intensive sectors of the American economy. Innovation occurs so rapidly that state-of-the-art equipment and designs are obsolete in less than five years. Given that the incentive to conduct R&D is obviously high, one would expect the resulting inventions to be well protected by patents. But until recently, most semiconductor designs could not be patented because they did not satisfy the test of nonobviousness applied by the courts. Other forms of intellectual property protection were even less effective. Trade secret protection is of limited use because much of the knowledge acquired in the development of a chip is evident in its layout. Employee turnover in the industry also contributes to the rapid diffusion of new techniques. And firms cannot invoke copyright protection because semiconductor designs are utilitarian in nature.[7]

[7] In 1976 Intel attempted to copyright a number of semiconductor chip "masks." These masks are stencils used to produce the various layers of a semiconductor chip via the process of photolithography. The Copyright Office refused to accept the masks. See Raskind (85), pp. 392-3.

Weak patent protection contributed to an industrywide tradition of "reverse engineering" competitors' products. Firms incorporate what they learn from their competitors' designs into new generations of their own. While the industry has long tolerated reverse engineering, concern over the protection of semiconductor designs increased as Japanese firms gained parity in manufacturing technology during the 1970s.[8] In 1979, and again in 1983, a number of semiconductor manufacturers lobbied Congress to extend copyright protection to semiconductor chip designs. These proposals failed because other firms objected to the complete elimination of reverse engineering implied by copyright protection. A compromise was achieved in the Semiconductor Chip Protection Act of 1984 (hereafter SCPA).[9]

The SCPA created a new form of intellectual property, called *mask works*. Owners of mask works are granted exclusive rights to reproduce, distribute, and import products embodying the mask work for a period of 10 years. Remedies for infringement include injunctions, lost profits, and sanctions up to $250,000. Protection does not extend to masks that are not original or that consist of commonplace, staple designs, or some combination of such designs that is not original.[10] This nonobviousness standard is much less stringent than what is required by patent law and somewhat more stringent than what is required for copyrights.[11]

---

[8] Indeed some observers believe that the American manufacturers moved to consolidate their comparative advantage, which was in R&D, when it became clear that they could no longer dominate Japanese firms on the basis of production technology alone. See Raskind and Stern (85), p. 263

[9] Pub. L. No. 98-620, tit. III, 17 U.S.C. 901-914 (Supp. II 1984).

[10] Raskind (85), p. 399.

[11] Stern (85), p. 318.

The primary limitation to the mask owner's exclusive rights is the defense of *reverse engineering*. The SCPA specifies a two-pronged test for the determination of infringement. The original and the allegedly infringing chip are compared. If they are "substantially identical," the court will find infringement. However, if the designs are only "substantially similar," the defendant is immune from liability if the following conditions are satisfied: (1) the defendant must demonstrate substantial toil and investment in the development of its chip; (2) the resulting design must satisfy the standard of originality specified by the SCPA.[12]

A successful defense of reverse engineering appears to require that the defendant's design be better than the plaintiff's. One author argues, "More likely, than not, for a defendant to prevail ... its resultant 'original mask work' must be one that is functionally superior to the protected work, as measured by the relevant technological criteria." He continues, "The proof of improvement, therefore, becomes the ultimate issue in establishing the defense of reverse engineering."[13]

Under certain, fairly weak conditions, the act allows a firm to appropriate another firm's technology without prior permission and without paying any royalty. This exception is unique to the SCPA. No other form of intellectual property protection links the questions of infringement to the obviousness/nonobviousness of a potentially infringing design. One expert, and a participant in the legislative process that produced the SCPA, states:

---

[12] The SCPA and its legislative histories do not provide a precise interpretation of the terms "substantially identical" and "substantially similar." Congress left these issues to be decided by the courts. To date there has been only one case, and the plaintiff prevailed. This ruling established that copying only part of a chip design could infringe a mask right, even where the defendant establishes extensive toil and effort. See Brooktree Corp. v. Advanced Micro Devices, Inc., 705 F. Supp. 491.

[13] Raskind (85), p. 402.

"Having any merit or qualifying for any kind of intellectual property protection is neither a necessary nor a sufficient condition to avoid infringement liability in patent or copyright law. An improvement patent is likely to infringe any 'dominant' patent to which it is 'subservient,' and a derivative work copyright often cannot be exploited without infringing the work from which it is derived. Section 906(a)(2) takes the unusual step of making this particular kind of derivative work, a reverse engineered mask work, free of subservience to the earlier work."[14]

While the SCPA increases the probability that semiconductor designs will be protected, the implied protection is of short duration. SCPA strikes a bargain between protecting the economic rents of existing inventions and encouraging firms to build on those discoveries as freely as possible. To do so, Congress specified a nonobviousness requirement considerably weaker than the one for patents.

B. The Federal Courts Improvement Act of 1982

The Federal Courts Improvement Act (FCIA) created a single venue for appeals of cases involving patents, trademarks, government contracts, tax, and international trade.[15] The new court, called the Court of Appeals for the Federal Circuit (CAFC), has been credited with increasing the

---

[14] Stern (85), p. 336.

[15] Pub. L. No 97-164.

uniformity of patent decisions.[16] We focus here on CAFC decisions that relaxed the effective standard of nonobviousness employed by the courts.[17]

Prior to passage of the FCIA, patents were regularly invalidated at trial, even though they had passed an investigation by the U.S. Patent and Trademarks Office.[18] CAFC decisions raised the presumption of patent validity and altered the test of nonobviousness.[19] Before the FCIA, the prevailing test of nonobviousness considered three factors: (1) the scope and content of the prior art; (2) the differences between the prior art and the patent claims; and (3) the level of ordinary skill in the relevant art.[20] Secondary factors, such as commercial success, failure of others, and long felt need, might also be relevant, but they could not override the three factors.[21] CAFC decisions elevated the secondary factors -- in particular, commercial success -- to a fourth and sometimes overriding factor.[22]

---

[16] See Sobel (88), p. 1090.

[17] Sobel (88) states, "The aggregate of the Federal Circuit's treatment of the obviousness issue of section 103 is probably the most important change it has made," p. 1091.

[18] One author finds that in the period 1953-77, 60% of patents adjudicated by the federal courts of appeals were invalidated. While these cases represent only a tiny proportion of all patent litigation in the period, the statistic is suggestive of the willingness of the courts to second guess the determinations of the Patent and Trademarks Office. See Szczepanski (87), p. 301, citing G. Koenig, *Patent Invalidity - A Statistical and Substantive Analysis,* 1980.

[19] Sobel (88), 1092-3. Prior to the FCIA, in some circumstances a party could prove an invention was obvious with merely "a preponderance of the evidence." The CAFC requires that an alleged defendant show "clear and convincing evidence" of patent invalidity.

[20] This test was developed in Graham v. John Deere, 383 U.S. 1, 17 (1966).

[21] Sobel (88), p. 1095.

[22] Ibid., p. 1095-6. In Stratoflex, Inc. v. Aeroquip Corp., 713 F.2d 1530 (Fed. Cir. 1983), the court stated: "Indeed, evidence of secondary considerations may often be the most probative and cogent evidence in the record. It may often establish that an invention appearing to have been obvious in light of the prior art was not."

What is the significance of these decisions? One author states, "Many patent attorneys believe that the obviousness defense is dead and that the cause of death lies in the decisions of the Court of Appeals for the Federal Circuit."[23] Another writes that "as a result of these changes, patents today are more likely to be held valid than, perhaps, at any time in our history."[24] An examination of the outcomes of subsequent appeals supports this claim.[25]

3. The Literature on Patent Attributes and Innovation

Research and development and intellectual property are subjects of considerable academic interest. This section reviews only work that is most closely related to this paper. The theory of optimal patents began with Nordhaus (69, 72) and Scherer (72), who examine how the competing objectives of providing an adequate incentive for R&D and minimizing monopoly distortions determines an optimal patent length. More recently, the work of Gilbert and Shapiro (90) and Klemperer (90) extend the analysis by allowing a social planner to vary patent *breadth* as well as length. Patent breadth represents the degree to which a product or process must differ from a patented one to avoid infringement of the patent. These models consider a single innovation and assume that the economic life of a patent is equivalent to the statutory length.

Patent breadth and nonobviousness are distinct characteristics. Patent breadth determines when the developer of a new invention must compensate the developer of a prior one. The

---

[23] Coolley (94), p. 625. The title of another paper is also suggestive: see Robert Desmond (93), "Nothing Seems Obvious to the Court of Appeals for the Federal Circuit..."

[24] Kastriner (91), p.23.

[25] Dunner (85) reports that the CAFC found 54% of the patents it reviewed to be nonobvious, as compared to a 30% rate in the initial trials. This compares favorably to the rates reported by Szczepanski (87) for previous years. The CAFC reversed 14% of lower court decisions that found an invention to be nonobvious. In contrast, it reversed 31% of lower court decisions that found an invention to be obvious. Coolley (89) reports that the CAFC reversed only 10% of cases that initially found the patent to be valid and 51% of cases where the patent was initially found to be invalid.

nonobviousness requirement distinguishes between proprietary and non-proprietary discoveries. An invention may be obvious and yet may not infringe an existing patent. Conversely, an invention may satisfy the standard of nonobviousness and yet still infringe the claims of a prior patent. The breadth of a patent depends on the nature of the invention and tends to be idiosyncratic.[26] In this paper, we focus exclusively on the standard of nonobviousness. To do so, we assume a system of patent rights, similar to the specifications of the Semiconductor Chip Protection Act, that eliminates the issue of infringement altogether.

A number of papers evaluate the role of patents in the context of cumulative discoveries, i.e., where inventions build on each other. Green and Scotchmer (92, 95) examine how patent breadth and length, licensing, and cooperative agreements determine the division of profits between two sequential innovators. Their definition of patent breadth is comparable to the test of nonobviousness employed in this paper. The interpretation, however, is quite different. They assume that both inventions are patentable and examine how patent breadth determines the division of proprietary rents between the initial and subsequent inventors. The model developed here abstracts from issues of licensing and the division of rents to focus on how the proportion of proprietary and non-proprietary discoveries affects the economic value of patents.

Scotchmer (96) considers the question of whether secondary inventions, made possible by an initial "pioneering" discovery, should be patentable. In a model where the initial inventor may contract with possible developers of derivative products, she shows that secondary inventions should

[26] In filing for a patent, an inventor lists one or more "claims" that represent the contribution of the invention over and above the prior art. The Patent and Trademarks Office examines, and possibly modifies, these claims before awarding the patent. Infringement is determined at trial by comparing the allegedly infringing product or process to the claims of the patent.

be unpatentable to ensure that the inventor of the pioneering discovery has the socially optimal incentive to conduct research. This result does not follow if ex-ante contracts are not feasible or when subsequent inventions do not depend entirely on the initial one.

Scotchmer and Green (90) show how the disclosure requirements of patent law may discourage firms from patenting intermediate discoveries, if by doing so they lose an advantage over their competitors in ongoing research. In a model of a two-stage patent race, they consider two standards of novelty (strong and weak) that really have the flavor of a nonobviousness requirement. They describe the circumstances when firms choose not to take advantage of the weaker standard.[27] In their model, a weaker nonobviousness requirement is never associated with less innovation. The model developed here, in which the horizon is infinite, shows that innovation may actually decrease because the weaker standard erodes the value of patents.

Merges (88, 92) argues that one role of patents and strict nonobviousness requirements is to encourage firms to engage in "risky" R&D projects, i.e., where there is less certainty of commercial success.[28] If less risky projects would be undertaken absent a patent, there is little social gain to extending protection to more obvious inventions. There is, however, the social cost of additional monopolies. There are additional losses if firms redirect their research toward less risky projects.

The R&D process used in this paper follows the initial models of stochastic patent races developed by Loury (79), Dasgupta and Stiglitz (80), and Lee and Wilde (80). These models consider an environment where several firms attempt to make a single discovery before their

---

[27] They also consider the incentive effects of the "first to file" and "first to invent rules," which are related to the issue of novelty.

[28] He also points out using evidence of commercial success as a test of nonobviousness may inadvertently reward better marketing and distribution rather than better products.

competitors. Reinganum (85) extends the Lee and Wilde (80) framework to a finite sequence of patent races. Each invention confers to its inventor a certain exogenous flow profit that continues until the next invention occurs. Each new discovery destroys the rents of the prior one, giving the model a flavor of Shumpeterian competition.

The model constructed here is similar in structure to Reinganum's (85), but differs in several ways. Flow profits depend on the extent of the innovation, which is stochastic. Not all inventions qualify for protection. An innovation failing to satisfy the nonobviousness requirement is appropriated by all firms, so it does not affect existing profits. The value of inventing, or failing to invent, the next improvement is endogenously determined and depends on the expected date of the next nonobvious invention.

Two recent papers report a result similar to the single peak result (Proposition 5) found here. O'Donoghue (98) constructs a model where firms choose their R&D intensity and the extent of their innovations with certainty. He shows that in the presence of transactions costs or costly monopoly distortions, a patent regime based on strict nonobviousness requirements is superior to a regime that requires innovators to license from prior inventors. Cadot and Lippman (95) construct a model of innovation where a technological leader engages in R&D competition with an imitator. The leader's incentive to innovate depends on the amount of time required for the imitator to reverse engineer the latest discovery. They show that the leader's R&D intensity is maximized by an intermediate delay between its discovery and successful imitation of the discovery.

The endogenous growth models of Aghion and Howitt (92) and Grossman and Helpman (91) describe a process of perpetual growth driven by a sequence of innovations generated by firms engaging in research and development. In a related paper, Lach and Rob (96) consider a single industry where firms purchase cost-reducing inventions and sink vintage-specific capital into production. We can interpret these models as an extreme case of the model constructed here, when all innovations satisfy the nonobviousness requirement and every discovery eliminates the rents associated with the prior one. By allowing for heterogeneity in the magnitude of discoveries, and conditioning patent protection on the basis of this magnitude, we can extend the analysis of these models by considering the effect of differing intellectual property regimes.

4. The Model

Time is continuous and the horizon is infinite. Discoveries occur at different points in time. It is convenient to divide time into the intervals between these discoveries. We call each of these intervals a patent race. During each patent race, firms compete to be the first to discover an invention. The race ends when a discovery occurs. The next race begins immediately after the discovery. Because there is randomness in the process that generates discoveries, the actual duration of patent races will vary.

A. Competition in R&D

There are n+1 firms, indexed by the superscript i. At the beginning of the race, firms simultaneously choose their R&D intensity, denoted $h_k^i \in \mathbb{R}$. Firms maintain their level of effort until a discovery occurs and the current race ends. The flow cost of conducting R&D, denoted $p \cdot C(h_k^i)$, is a strictly increasing, twice continuously differentiable function of R&D intensity. The coefficient p represents the relative price of R&D inputs. All firms share the same R&D technology.

A firm's R&D intensity affects its rate of discovery, but does not determine the exact date that it will make a discovery. Instead, R&D intensity determines a probability distribution over invention dates. On average, a firm that exerts more effort will make a discovery before a firm that exerts less effort.[29] A simple way to capture these properties is to assume that discoveries arrive through time according to a Poisson process, where the arrival rate of discoveries is determined by the R&D intensity of the firm. We assume that the arrival rate of ideas for firm *i* is $\lambda \cdot h_k^i$, where λ represents an industry-specific productivity parameter.

Let t_k^i denote the time elapsed in the kth patent race before firm *i* makes a discovery. If firm *i* chooses the R&D intensity h_k^i, the probability that it discovers an invention before date t is

$$Pr\{t_k^i < t\} = 1 - e^{-\lambda h_k^i \cdot t}.$$

One characteristic of Poisson processes is that they are *memoryless*. This means that the *hazard rate* of discoveries -- the probability that a firm makes a discovery in the next instant of time, given that it has not already made a discovery -- is constant and equal to the firm's R&D intensity. This is also true for the firm's competitors, so it faces a constant *rival hazard rate* $\lambda a_k^i \equiv \lambda \cdot \sum_{j \neq i} h_k^i$ during the kth race. Firm *i* wins the race if it invents before any other firm. The probability that it wins is $\Pr\{t_k^i \le t_k^j,\ \forall j \neq i\} = h_k^i / [h_k^i + a_k^i]$, the ratio of firm *i* 's hazard rate to the hazard rate for the entire industry.

We make one further assumption about the nature of technological competition: Firms that discover a patentable invention do not compete in the subsequent patent race.[30] A firm that owns a patented invention will be called an *incumbent*. The other firms will be called *challengers*.

---

[29] A firm's choice of R&D spending affects the distribution over invention dates in the sense of first order stochastic dominance.

[30] While this is an ad hoc restriction on behavior, it considerably simplifies the model and subsequent analysis. In other models, the incumbent can be shown to compete less aggressively, or not at all. See Grossman and Helpman (91) and Reinganum (85).

B. Inventions and the System of Property Rights

A discovery is an improvement in product quality. The extent of an improvement is denoted $u_k \in [0, \bar{u}]$, $\bar{u} < \infty$.[31] The magnitude of improvements is random, unknown until the time of invention, and common knowledge thereafter. For each invention, u is drawn from the continuous density $f(u)$ with corresponding cumulative density $F(u)$. This distribution is constant through time and unaffected by the level of a firm's R&D spending.[32]

We assume that once a discovery has been made, it can be reverse-engineered at zero cost by all other firms. This means that discoveries are only proprietary if they are protected by the system of intellectual property rights. If a patent is granted, the inventor receives an exclusive right to produce and sell that invention. For simplicity, we assume that the statutory life of the patent is infinite. Not all inventions will be protected, however. Let $s \in [0, \bar{u}]$ denote the minimum extent of improvement for which the patent office is willing to grant a patent. This represents the standard of nonobviousness. An invention whose extent is less than s is not protected and becomes part of the public domain of product improvements. Let $\theta(s) = 1 - F(s)$ denote the ex-ante probability of obtaining patent protection, given the standard of nonobviousness s.

Patent claims are defined as the improvement itself, so that each improvement does not infringe the patent on another. We must make some assumption about the right of an inventor to use

---

[31] Alternatively, we can express innovations as some percent reduction in the cost of producing some final good. In this case $\bar{u} < 1$. The analysis will be consistent with the quality improvement representation if we assume that cost reductions are perfectly compatible, so that a cost reduction applied to different vintages of technology achieves the same percent reduction in cost.

[32] More generally, we expect that firms affect the expected date *and* magnitude of their discoveries. Although we ignore the latter from the analysis, the nonobviousness requirement may indeed affect the ambitiousness of R&D programs. See Merges (88).

prior generations of improvements. For example, firms may be required to license all prior improvements from their inventors. At the other extreme, we could assume that an inventor may use all prior discoveries without obtaining a license. We assume an intermediate case: if an invention satisfies the standard of nonobviousness, the inventor may use all prior discoveries without licensing them.[33] However, if the standard is not satisfied, the prior discoveries remain proprietary.

One implication of this specification is that there is always, at most, one protected invention. Each time another patentable discovery is made, the inventor of the last patented invention loses her exclusive rights. Thus while the statutory length of patent protection is infinite, the economic life of a patent is the amount of time until the next patentable invention.

While this is an arbitrary definition of property rights, it captures in a very tractable way the phenomenon we wish to study. We focus on how the accumulation of proprietary inventions erodes the profits associated with an existing patent. The simplest way to represent this is to assume that the first proprietary invention eliminates those rents entirely, while nonproprietary inventions have no effect at all.[34] The definition reflects the system of property rights specified by the Semiconductor Chip Protection Act of 1984 (SCPA).[35] Under the SCPA, a firm that reverse engineers another firm's

[33] We are not concerned here with how firms might share the rents of sequential innovations. That question is addressed in Scotchmer (96) and Green and Scotchmer (92, 95).

[34] Lach and Rob (96) use a more natural approach by combining new technology with the acquisition of vintage-specific production equipment. In a model of Cournot competition, the introduction of new technologies leads to a more gradual erosion of profits until the older firms exit altogether.

[35] Section 2 of the paper describes this law in greater detail.

product need not pay royalties if the design it develops is sufficiently better than the original and the standard of sufficiently better is the standard of nonobviousness.[36]

C. The Output Market and Flow Profits

All consumers are identical and aggregate demand is normalized to one. The reservation value of the final product to consumers is simply the level of its quality.[37] The *best available technology* during the kth patent race is a product embodying all of the improvements that have already occurred. We denote the associated reservation value as $U_k \equiv \sum_{y=1}^{k-1} u_y$. Let u_k^p denote the extent of the innovation protected during the kth race.

Firms compete in prices and the cost of production is zero. The system of property rights described above implies that the incumbent can offer the best available technology U_k while other firms can offer $U_k - u_k^p$. Then, the equilibrium price of the final good is u_k^p, the incumbent earns flow profit u_k^p, and all challengers earn a flow profit of zero.

---

36 This description is also a natural generalization of the property rights assumed in Reinganum (85), Grossman and Helpman (91), Lach and Rob (96), and Aghion and Howitt (92). In each of these models, every innovation receives protection and the length of this protection is only the amount of time until the next invention. This is equivalent to setting the standard of nonobviousness in this model to $s = 0$.

37 These assumptions make the consumer's problem stationary. If we characterize innovations as cost reductions, we get the same behavior by assuming a constant elasticity of demand function with an elasticity of one.

Table 1: Flow profits earned during the *k*+1st race as a function of the kth discovery

	The *k*th invention is	
The firm earns	Patentable	Unpatentable
Incumbent of the *k*th race	0	u_k^p
Winning challenger	u_k	0
Losing challenger	0	0

Firm *i*'s flow profits during the k+1st race depend on two outcomes that occur in the kth race. The first is the determination of which firm invents first. The second is the determination of the invention's magnitude. If firm *i* discovers an improvement and it satisfies the nonobviousness standard, its flow profits during the k+1st race are u_k. If the improvement satisfies the nonobviousness standard, but is discovered by another firm, *i*'s flow profits will be zero. If the improvement is found to be obvious, the flow profits of each firm in the next race will be the same as those earned in the current race.

5. Equilibrium

This section establishes the existence and properties of a stationary, symmetric equilibrium of the game described in the previous section. During each patent race, challengers choose a level of R&D activity that balances spending in the current period against the expected gains earned in future races. The expected gains are a weighted average of the values of winning and losing the current race, discounted to reflect the expected length of the current race. The choice of R&D intensity affects the probabilities of winning and losing the current race, as well as the expected length of the current race.

A. Constructing the Firm's Problem

During each race, firms choose the R&D intensity that maximizes the expected present value of current cash flow plus the expected present value of competing optimally in future races. For challengers, current cash flows are R&D expenditures, $-p{\cdot}C(h_k^i)$. The value of competing in the next race is a weighted average of the discounted values of the position conferred by winning or losing the current race.

We use V^w to denote the expected value of playing optimally in the next race when the firm wins the current race. We use V^l to denote the expected value of playing optimally in the next race when the firm loses the current race. We will refer to these as *continuation* values. These values could vary across races, but not in the stationary equilibrium we study, so we suppress the time subscripts on these values. In the stationary equilibrium we examine, the behavior of firms in the current race does not affect the continuation values. For the moment, we will treat V^w and V^l as parameters. Later, we will compute their values in equilibrium.

To compute the weighted averages of the possible outcomes of the current race, we introduce the following probabilities:

i) The probability that a firm makes a discovery before date t is $(1-e^{-\lambda \cdot h_k^i \cdot t})$;

ii) The probability that any of a firm's rivals makes a discovery before date t is $(1-e^{-\lambda \cdot a_k^i \cdot t})$;

iii) The probability a firm has not made a discovery by date t is $e^{-\lambda \cdot h_k^i \cdot t}$;

iv) The probability that no firm has made a discovery by date t is $e^{-\lambda\cdot(h_k^i+a_k^i)\cdot t}$;

v) The probability that firm i makes a discovery *at* date t, given that it hasn't yet made a discovery is $\lambda\cdot h_k^i\cdot e^{-\lambda\cdot h_k^i\cdot t}$.

Firms incur R&D expenses until the first discovery occurs. The present value of R&D spending depends on the expected date of that discovery and the discount rate, denoted r:

$$\int_0^\infty -pC(h_k^i)\cdot e^{-\lambda\cdot(h_k^i+a_k^i)\cdot t}\cdot e^{-r\cdot t}dt = \frac{-pC(h_k^i)}{\lambda\cdot(h_k^i+a_k^i)+r}.$$

The fraction $1/[\lambda(h_k^i+a_k^i)+r]$ is the discount factor that converts flow R&D spending into the expected present value of R&D expenditures over the current race. It is a measure of the expected length of the current race. If the industry's R&D intensity is low, the expected date of the next discovery is far off. The discount factor is relatively large, reflecting the fact that R&D spending will probably go on for some time. If the industry's R&D intensity is high, the discount factor is smaller, because the next discovery is likely to occur very soon.

When the race ends, each firm is either a winner or a loser. The expected present value of winning the race is

$$\int_0^\infty V^w\cdot\lambda h_k^i\cdot e^{-\lambda\cdot(h_k^i+a_k^i)\cdot t}\cdot e^{-r\cdot t}dt = \frac{\lambda h_k^i\cdot V^w}{\lambda\cdot(h_k^i+a_k^i)+r}$$

which depends on the expected date of the first discovery and the probability of being the inventor of that discovery. Similarly, the expected present value of losing the current race is

$$\int_0^\infty V^l \cdot \lambda a_k^i \cdot e^{-\lambda \cdot (h_k^i + a_k^i) \cdot t} \cdot e^{-r \cdot t} dt = \frac{\lambda a_k^i \cdot V^l}{\lambda \cdot (h_k^i + a_k^i) + r}.$$

The firms' objective function, denoted $V^i(h_k^i, a_k^i)$, is the sum of the three preceding terms:

$$V^i(h_k^i, a_k^i) = \frac{\lambda [h_k^i \cdot V^w + a_k^i \cdot V^l] - pC(h_k^i)}{\lambda \cdot (h_k^i + a_k^i) + r}. \qquad [1]$$

We make the following assumptions about the R&D technology:

(i) $h \in [0, \bar{h}]$ s.t. $\bar{h}$, $C(\bar{h}) < \infty$;

(ii) $C'(h) > 0$, $C''(h) > 0$, $\forall\, h > 0$;

(iii) $\lim_{h \to 0} C(h)/h = \lim_{h \to 0} C'(h) = 0$;

(iv) $\lim_{\bar{h} \to \infty} C'(\bar{h}) = \infty$;

(v) $C'(h)h - C(h) \geq 0$;

(vi) $C''(h)h - C'(h) \geq 0$;

(vii) $\lambda(n+1) \geq 1$.

The first assumption states that R&D programs are subject to saturation.[38] The second states that R&D effort is costly at the margin and the marginal cost of additional effort is increasing. The third and fourth assumptions ensure that the firm's objective function is maximized by some intermediate level of R&D effort. The fifth assumption implies that production of R&D intensity is subject to diminishing returns. The sixth assumption states that the R&D cost function is sufficiently convex.[39] The seventh assumption requires that the R&D productivity coefficient not be too small.[40]

---

[38] This assumption guarantees that firms' current returns are bounded.

[39] A quadratic cost function, for example, satisfies this condition.

[40] Assumptions (vi) and (vii) are stability conditions that determine a number of properties of the equilibrium. Assumption (vii) is also a necessary condition for the uniqueness of the equilibrium.

The first order condition of the firms' problem identifies the level of R&D intensity where the marginal cost of additional effort is just equal to the expected gain associated with winning the current race:

$$pC'(h_k^i) = \lambda\left[V^w - V^i(h_k^i, a_k^i)\right]. \qquad [2]$$

Substituting for $V^i(h_k^i, a_k^i)$ in [2], the first order condition can also be expressed as

$$pC'(h_k^i) = \lambda\left(\frac{rV^w + pC(h_k^i)}{\lambda(h_k^i + a_k^i) + r} + \frac{\lambda a_k^i [V^w - V^l]}{\lambda(h_k^i + a_k^i) + r}\right). \qquad [3]$$

Equation [3] relates the marginal cost of additional R&D to the sum of two present values. The first term is called the *replacement* effect because it measures the value of replacing ongoing R&D expenditures with the continuation value of being the winner. The second term is called the *rivalry* effect because it measures the difference between the values of starting the next race as the winner or loser of the current one. Firms will spend more on R&D when either of these two values increases. The replacement effect is more important when there is little effective rivalry (λa_k^i is small). Then the firm worries less about losing the race and more about how soon it can replace its current expenditures with flow profits. When the firm encounters more effective rivalry (λa_k^i is large), the rivalry effect is more important than the replacement effect. In this case, the firm worries more about losing the current race than about how soon it can replace its current R&D spending.

B. The Stationary Symmetric Equilibrium

A strategy of a firm in the game is a specification of a feasible R&D intensity to be played in each race, for each possible history of the game preceding that race. When the firm is the incumbent, its only feasible R&D intensity is zero. Whenever the firm is a challenger, the set of feasible R&D intensities is always the same subset of $\mathbb{R}$. Payoffs are the discounted sum of the present values of any flow profits or R&D expenditures that occur in the patent races. At the beginning of each race, each firm knows the play of all firms in the prior races and the outcomes of those races. There are likely to be many equilibria of the game, but we choose to focus on stationary equilibria where firms choose identical strategies. We can show the following:

Proposition 1: If the R&D cost function satisfies the assumptions (i)-(vii) above, there exists a unique stationary symmetric equilibrium of the game.

In the appendix, we show that assumptions (ii) - (v) establish the existence and uniqueness of a best response, for a given level of rivalry and finite continuation values $V^w \geq V^l$, in individual stage games. This best response is described by the first order condition to the firm's problem, i.e., [3]. We then compute the equation that describes the best response of firms when they all choose the same R&D intensity and provide a sufficient condition for the uniqueness of the symmetric best response in the stage games.

The structure of the game is stationary in the sense that the R&D production technology, the distribution over invention magnitudes, and the relationship between patented technology and resulting profits do not vary across races. The expected outcome of the races, then, varies only if firms choose different R&D intensities over time. We will examine an equilibrium where firms respond to the same conditions in the same way through time.

The exact magnitude of flow profits associated with a patentable discovery is not known until the discovery has actually occurred. Firms use the expected value of flow profits, denoted u^e, when choosing their R&D intensity:

$$u^e \equiv u^e(s) = \int_s^{\bar{u}} u \cdot f(u) / [1 - F(s)] du,$$

where $f(u)/[1-F(s)] = f(u)/\theta$ is the conditional probability distribution over patentable discoveries.

The incumbent of the next race earns flow profits for the duration of that race. When that race ends, one of two things will happen. If the next race results in a patentable discovery, the incumbent's flow profits are eliminated and it will begin the subsequent race as a challenger. But if the next race is ended by an unpatentable discovery, the incumbent continues to earn its flow profit during the subsequent race. As long as the firm is the incumbent, it faces the same set of possible outcomes in the next race.

A firm that begins the next race as a challenger spends a constant flow cost for the duration of the race. If it loses that race, it begins the subsequent race as a challenger. If it wins the next race, and its discovery is patentable, it begins the subsequent race as an incumbent. However, if its discovery is not patentable, it continues as a challenger in the subsequent race. As long as the firm is a challenger, it faces the same set of possible outcomes in the next race.

Suppose that all firms choose the same R&D intensity, *h*, in all future races. The probabilities of winning and losing, together with the expected length of races, will be the same in each race. Ex ante, each race looks like any other. The values of being an incumbent or a challenger, denoted $V^I(h)$ and $V^C(h)$, respectively, also do not change. The only question is whether a firm begins the race as the incumbent or as a challenger. The expected value of losing is the expected value of being a challenger in the subsequent race, i.e., $V^l = V^C(h)$. The expected value of winning is not, however, the value of being the incumbent in the next race. If the resulting discovery is patentable, the firm indeed enjoys the expected value of starting the next race as the incumbent.

But if the discovery is unpatentable, the firm begins the next race as a challenger. The expected value of winning the current race is then a weighted average of the values of being an incumbent or a challenger in the next race, i.e., $V^{w} = \theta V^{I}(h) + (1-\theta)V^{C}(h)$. The difference in the values of winning and losing a race is then $[V^{w} - V^{l}] = \theta[V^{I}(h) - V^{C}(h)]$, the difference between the values of being the incumbent and a challenger times the probability that a firm's discovery is patentable.

This structure implies that the expected value of being an incumbent in the next race is

$$V^{I}(h) = \frac{u^{e} + \theta\lambda nhV^{C}(h)}{\lambda nh + r} + \frac{(1-\theta)\lambda nh}{\lambda nh + r}\cdot\left[\frac{u^{e} + \theta\lambda nhV^{C}(h)}{\lambda nh + r} + \frac{(1-\theta)\lambda nh}{\lambda nh + r}\cdot\left[\ \dots\right.\right. \text{ or}$$

$$V^{I}(h) = \left(\frac{u^{e} + \theta\lambda nhV^{C}(h)}{\lambda nh + r}\right)\cdot\left[1 + \sum_{x=1}^{\infty}\left(\frac{(1-\theta)\lambda nh}{\lambda nh + r}\right)^{x}\right] = \left(\frac{u^{e} + \theta\lambda nhV^{C}(h)}{\theta\lambda nh + r}\right).$$

Similarly the expected value of starting the next race as a challenger is

$$V^{C}(h) = \left(\frac{\theta\lambda h\left[V^{I}(h) + (n-1)V^{C}(h)\right] - pC(h)}{\lambda nh + r}\right)\cdot\left[1 + \sum_{x=1}^{\infty}\left(\frac{(1-\theta)\lambda nh}{\lambda nh + r}\right)^{x}\right] = \left(\frac{\theta\lambda hV^{I}(h) - pC(h)}{\theta\lambda h + r}\right).$$

Because firms switch between being incumbents and challengers, the continuation values $V^{I}(h)$ and $V^{C}(h)$ are functions of each other. Using the preceding equations, we find that

$$V^{I}(h) = \frac{\left[r + \theta\lambda h\right]\cdot u^{e} - \theta\lambda nh\cdot pC(h)}{r[r + \theta\lambda(n+1)h]}, \qquad [4a] \qquad\qquad V^{C}(h) = \frac{\theta\lambda h\cdot u^{e} - \left[r + \theta\lambda nh\right]\cdot pC(h)}{r[r + \theta\lambda(n+1)h]}. \qquad [4b]$$

The stationary behavior of firms implies continuation values that are a weighted average of the expected flow profit enjoyed by incumbents and the R&D expenditures of challengers. On average, firms spend $\theta\lambda h/[r+\theta\lambda(n+1)h]$ of time as incumbents. On average, firms spend $\theta\lambda nh/[r+\theta\lambda(n+1)h]$ of time as challengers. These are the stationary (transition) probabilities of moving from one state (incumbent or challenger) to the other. The remaining fraction of time,

$r/[r+\theta\lambda(n+1)h]$, accounts for the proportion of time that firms remain in their initial state. Multiplying the associated flow profit and costs by these probabilities generates weighted averages of the expected cash flow a firm will earn over all future patent races. Dividing by r converts the average flows into present values.

But $\theta\lambda h$ is also the probability that a given firm will make a patentable discovery in the next instant of time. The average amount of time between discoveries is $1/\theta\lambda(n+1)h$. Incorporating the discount rate r, we can convert the cash flow earned over the next race into present values by using the coefficient $1/[r+\theta\lambda(n+1)h]$, the discount factor associated with the average length of patent races. Thus when we consider the difference in the values of being an incumbent or a challenger in the next race, we find that

$$[V^I(h) - V^C(h)] = \frac{u^e + pC(h)}{r + \theta\lambda(n+1)h},$$

which is the present value of earning the incumbent's expected flow profit and avoiding a challenger's R&D expenditures over the expected length of a race. The denominator is also a measure of the *economic* life of patents. As $\theta\lambda(n+1)h$ becomes larger, patentable discoveries occur more frequently. The incumbent enjoys its gain for less time, on average, so the present value of the gain is smaller.

Because $V^l = V^C(h)$ and $V^w = \theta V^I(h) + (1-\theta)V^C(h)$, we can define the symmetric best response of firms as a function of V^I and V^C. A symmetric stationary equilibrium of the game is an R&D intensity, denoted σ, that is a best response to the continuation values $V^I(\sigma)$ and $V^C(\sigma)$. In the appendix, we show that equations [3], [4a], and [4b] imply that σ is the solution to the following equation:

$$pC'(h) = \theta \cdot [V^I(h) - V^C(h)] + (1-\theta) \cdot 0. \qquad [5]$$

The left hand side of [5] is the cost of additional R&D effort. At the margin, this is the cost of moving the firm's expected invention date just ahead of any of its competitors. The right hand side is the expected gain associated with becoming the first inventor, taking into account the probability that the invention is patentable. The gain is measured relative to the expected value of being a challenger in the next race. Thus, if the invention is obvious, a challenger remains a challenger and gains nothing. The larger the expected gain associated with inventing first, the more firms are willing to spend to invent faster.

C. Properties of the Equilibrium

The model allows us to differentiate industries by the number of firms, R&D productivity, discount rates, and R&D input prices. Each of these parameters affects the equilibrium R&D intensity of firms. We can show the following:

Proposition 2 - If the R&D technology satisfies (vi) and (vii), the industry rate of innovation is increasing in the number of firms and the productivity of R&D, but decreasing in the discount rate and the price of R&D inputs.

Increasing the number of firms or the productivity of R&D activity causes per firm R&D activity to *decline*. But Proposition 2 states that such changes increase the *industrywide* rate of innovation when the cost function for R&D is sufficiently convex (assumption vi) and the productivity parameter is not too small (assumption vii). These assumptions can be thought of as stability conditions because they imply that firms' responses to changes in the parameters are not too great.[41] In the appendix, we show that firms' R&D activity is inelastic with respect to changes in the number of firms or the productivity of R&D. Thus while firms decrease their activity, the decline is not sufficient to offset the activity of the additional firm or the additional productivity of R&D.

As the discount rate increases, firms discount future gains more heavily, reducing their value. The equilibrium condition states that the marginal cost of current R&D equals the expected gain. Because the expected gain declines in value, firms reduce their R&D activity. Similarly, increasing the R&D input price causes the marginal cost of current R&D efforts to rise. Firms compensate by reducing their R&D activity until the marginal cost and the expected gain are again equal. In the appendix, we show that, under assumptions vi and vii, firms' R&D activity is inelastic with respect to changes in input prices. Thus while activity declines, R&D expenditures will increase.

The model suggests that industries with more firms, more productive R&D, lower discount rates, and lower input prices will innovate more rapidly. A less obvious implication of the preceding analysis is that the opportunity cost of increasing R&D at the firm level is lower for firms in rapidly innovating industries. For example, in industries with more firms, lower input prices, or higher productivity, the marginal cost of R&D is lower. And while the marginal cost of R&D in industries with lower discount rates is higher, each dollar of R&D is associated with a larger expected gain associated with inventing first. This intuition will be important when we analyze the effects of changes in the nonobviousness requirement on different industries.

---

[41] Other racing models, including Loury [79], Lee and Wilde [80], and Reinganum [85,] assume that similar stability conditions are satisfied.

6. Changing the Nonobviousness Requirement

A. The General Problem

Suppose we make the nonobviousness requirement more strict (increase *s*).[42] How will firms respond? We know that in equilibrium firms equate the marginal cost of additional R&D effort to the expected gain associated with inventing first. If stricter nonobviousness requirements cause the expected gain to rise, firms will increase their R&D activity. If the expected gain declines, firms will reduce their R&D activity.

Increasing the standard of nonobviousness causes the expected gain to inventing first to change in more than one way. It is useful to separate these changes into what we call *static* and *dynamic* effects:

$$\frac{d[V^I - V^C]}{ds} = [V^I - V^C] \cdot \frac{\partial \theta}{\partial s} + \theta \cdot \left(\frac{\partial V^I}{\partial s} - \frac{\partial V^C}{\partial s} \right). \qquad [6]$$

The static effect of an increase in the nonobviousness requirement is the change in the probability that a firm's next discovery will be patentable. This is the first term on the right hand side of equation [6]. The dynamic effect of stricter nonobviousness requirements is the change in the values of being the incumbent and a challenger in the subsequent patent races. This is captured in the second term of [6].

Stricter nonobviousness requirements reduce the likelihood that inventions qualify for patent protection ($\partial\theta/\partial s < 0$), so the static effect is always negative. Conversely, relaxing the nonobviousness requirement increases the probability that a firm's next discovery will be patentable

---

[42] To avoid confusion over signs, the following calculations assume an *increase* in the nonobviousness requirement.

and thus increases the expected gain to inventing first. This should encourage firms to conduct more R&D. This is precisely the intuition behind many of the proposals to relax patentability standards in the 1980s.

But the analysis is not complete until we consider the direction and magnitude of the dynamic effect. It is the difference in the effects on the continuation values for the incumbent and challengers. We can show that

$$\frac{\partial V^I(\sigma)}{\partial s} = \frac{f(s)}{\theta} \cdot \left\{ \left(\frac{r+\theta\lambda\sigma}{r+\theta\lambda(n+1)\sigma} \right) \cdot \frac{[u^e(s)-s]}{r} + \left(\frac{\theta\lambda n\sigma}{r+\theta\lambda(n+1)\sigma} \right) \cdot [V^I(\sigma)-V^C(\sigma)] \right\}, \qquad [7]$$

and

$$\frac{\partial V^C(\sigma)}{\partial s} = \frac{f(s)}{\theta} \cdot \left(\frac{\theta\lambda\sigma}{r+\theta\lambda(n+1)\sigma} \right) \cdot \left\{ \frac{[u^e(s)-s]}{r} - [V^I(\sigma)-V^C(\sigma)] \right\}. \qquad [8]$$

For the incumbent, there are two benefits of stricter nonobviousness requirements. First, the average flow profit of patentable discoveries increases.[43] The present value of this increase is $[u^e(s)-s]/r$. The incumbent enjoys this improvement $[r+\theta\lambda\sigma]/[r+\theta\lambda(n+1)\sigma]$ percent of the time. This is the first term in brackets in [7]. Second, because fewer discoveries qualify for protection, a smaller proportion of future discoveries will eliminate the incumbent's patent. This increases the average amount of time between patentable discoveries and thus the average length of patent races. In other words, the economic life of patents has increased. The gain associated with longer patent life is the difference in the values of being an incumbent and a challenger, multiplied by the initial proportion of time the incumbent would have spent as a challenger, $\theta\lambda n\sigma/[r+\theta\lambda(n+1)\sigma]$. This is the second

---

[43] Note that the derivatives taken in equations [7] and [8] are done holding constant the level of future R&D.

term in [7]. The second term can also be interpreted as the present value of remaining the incumbent, discounted to reflect the expected date on which the incumbent's patent would have been eliminated by a marginal discovery. This is more important when the economic life of patents is short, because the benefit is discounted less heavily than when patents are expected to last a long time.

Equation [8] shows that there is a benefit and a cost of stricter nonobviousness requirements for challengers. The benefit is again the increase in the average flow profit of patentable discoveries, which a challenger enjoys $\theta\lambda\sigma/[r+\theta\lambda(n+1)\sigma]$ percent of the time. The cost is the fact that a challenger's next discovery is less likely to be patentable. After the change in standards, when a challenger makes a marginal discovery ($u = s$), it remains a challenger. The loss is the difference in the values of being an incumbent and a challenger, multiplied by the initial proportion of the time it could expect to be an incumbent, $\theta\lambda\sigma/[r+\theta\lambda(n+1)\sigma]$. The cost can also be interpreted as the present value of the loss associated with the increase in the average length of patent races. The loss is discounted to reflect the expected date on which the challenger would have made a patentable discovery under the old standard.

Equations [7] and [8] show that stricter nonobviousness requirements raise the value of being an incumbent and may increase or decrease the value of being a challenger. But the net effect of these changes is unambiguously positive:

$$\frac{\partial V^I}{\partial s} - \frac{\partial V^C}{\partial s} = \frac{f(s)}{\theta} \cdot \left\{ \left(\frac{r}{r+\theta\lambda(n+1)\sigma} \right) \cdot \frac{[u^e(s)-s]}{r} + \left(\frac{\theta\lambda(n+1)\sigma}{r+\theta\lambda(n+1)\sigma} \right) \cdot [V^I(\sigma) - V^C(\sigma)] \right\} > 0. \qquad [9]$$

The dynamic effect is a weighted average of the benefits associated with the higher average flow profit of patentable discoveries and the longer economic life of patents. The weights depend on $\theta\lambda(n+1)\sigma$, the industrywide arrival rate of patentable discoveries. The average length of time between patentable discoveries, and thus the economic life of patents, is inversely related to this arrival rate. When patentable discoveries are infrequent, that is, when $\theta\lambda(n+1)\sigma$ is small, the increase in their average profitability is relatively more important than the increase in the economic life of patents. When patentable discoveries occur frequently, the increase in the economic life of patents is relatively more important than the increase in their average profitability.

It is now clear that the static and dynamic effects of stricter nonobviousness requirements work in opposite directions. Tighter patentability standards decrease the likelihood that a firm's next discovery will be patentable, which reduces the expected gain associated with R&D. On the other hand, tighter standards increase the expected value of being an incumbent relative to the expected value of being a challenger. This raises the expected gain associated with R&D. This ambiguity raises two questions. First, does one of these effects always dominate? If the static effect is always stronger, the conventional wisdom motivating many recent changes in intellectual property law is essentially sound. But if the static effect doesn't always dominate the dynamic effect, can we say anything about when one effect is more important than the other? These questions are addressed in the following sections.

B. The Relative Strength of the Static and Dynamic Effects

The response of firms to a change in the nonobviousness requirement depends on its effect on the expected gain to inventing first. Suppose the standard of nonobviousness is increased. The expected gain to inventing first is increased by a rise in the average profitability of patented inventions but is also reduced by the lost cash flow of marginal discoveries that are no longer patentable. Combining equations [6] and [9], we find

Proposition 3: The static effect dominates the dynamic effect when the following expression is negative:

$$\Psi(s) \equiv \left(\frac{\theta\lambda(n+1)\sigma}{r+\theta\lambda(n+1)\sigma} \right) \cdot [u^{e}(s) - s] - \left(\frac{r}{r+\theta\lambda(n+1)\sigma} \right) \cdot [s + pC(\sigma)].$$

The dynamic effect dominates the static effect when $\Psi(s) > 0$.

Corollary 4: Relaxing the nonobviousness requirement will increase R&D activity when the static effect dominates the dynamic effect, i.e., when $\Psi(s) < 0$. It will decrease R&D activity when $\Psi(s) > 0$.

As the nonobviousness requirement is made more strict, firms encounter the following tradeoff. On the one hand, a firm that makes a marginal discovery fails to obtain a patent and continues as a challenger in the next race. It loses the associated profit and the cost of R&D spending it would have avoided for the length of the next race. The present value of the lost cash flow is increasing in the expected length of the next race. When patentable discoveries are infrequent, these losses are relatively large. But when patentable discoveries occur frequently, the value of these losses is smaller. On the other hand, a stricter nonobviousness requirement raises the average flow profit of patentable discoveries. The associated gain is increasing in the frequency of patentable discoveries. The net effect is a weighted average of these cash flow gains and losses, where the weights are determined by the industrywide arrival rate of patentable discoveries. This is the intuition of Proposition 3.

Corollary 4 uses Proposition 3 to address the policy question raised in this paper. If the weighted average computed above is negative, reductions in the standard of nonobviousness will indeed increase firms' R&D activity and consequently the rate of innovation. When this weighted average is positive, R&D activity and the rate of innovation will decline.

We have not yet established whether the dynamic effect is ever more important than the static effect. We do this in the following proposition:

Proposition 5: There is a unique standard of nonobviousness, denoted s^*, such that in the interval $[0, s^*)$, R&D activity is strictly *increasing* in the standard of nonobviousness.

In the appendix, we show that when the standard of nonobviousness is very weak, the dynamic effect is stronger than the static effect. Consequently, increases in the standard raise R&D activity. As we make the requirement more strict, the dynamic effect becomes smaller while the static effect becomes larger. When the nonobviousness requirement is very strict, the static effect dominates. There is only one standard of nonobviousness, where the two effects are exactly equal.

We can use the critical nonobviousness standard, s^*, to divide the interval of possible nonobviousness requirements $[0, \bar{u}]$ into two regions. In the interval $[0, s^*)$, an increase in the standard raises R&D activity. We call this the *dynamic region*. In the interval $(s^*, \bar{u}]$, an increase in the standard causes R&D activity to fall. We call this the *static* region. R&D activity and the rate of innovation is maximized by setting the nonobviousness requirement at s^* and protecting $[1-F(s^*)]$ percent of all discoveries.[44]

[44] Throughout this paper, we implicitly assume that increasing R&D activity is socially beneficial. This corresponds with the sentiments of the proponents for intellectual property reform in the 1980s. It is possible, however, that firms will overinvest in R&D relative to the socially optimal level because the patent system rewards only firms that invent first. This is often referred to as the *racin*g effect and is discussed extensively in the industrial organization literature.

C. Implications for Rapidly Innovating Industries

In this section, we investigate the factors that determine the relative size of the dynamic and static regions identified in the preceding section. The size of these regions varies with the factors that affect the equilibrium rate of innovation in an industry. We can use these relationships to tentatively evaluate some of the changes in intellectual property law introduced during the 1980s.

Recall that the dynamic effect is stronger than the static effect when $\Psi(s) > 0$. Greater weight is placed on the increase in average profitability of patentable discoveries when the industrywide arrival rate of patentable discoveries is high. This suggests, but does not prove, that the dynamic effect is more important in industries that innovate rapidly. There is potential for ambiguity if the flow R&D expenditures of firms in such industries are higher. Higher R&D spending makes the second term more negative, causing $\Psi(s)$ to fall. In the appendix, we prove the following:

Proposition 6: The critical standard of nonobviousness, s^*, is increasing in the equilibrium rate of innovation.

Corollary 7: In rapidly innovating industries, a smaller proportion of inventions can be protected by patents without causing the rate of innovation to decline.

In section 5C, we showed that lower input prices, higher R&D productivity, more firms, or lower discount rates were associated with higher rates of innovation, but not necessarily higher levels of per firm R&D spending. Increases in R&D productivity or the number of firms actually reduce per firm R&D activity and expenditures. Decreases in the discount rate increase per firm R&D spending but also shift the probability weight away from the cash flow losses. Lower input prices induce higher R&D activity, but not so much that actual expenditures rise. For each of the parameters studied, firms in rapidly innovating industries face a lower opportunity cost to raising their R&D activity than do firms in industries that innovate more slowly. Consequently, the dynamic effect dominates over a larger range of nonobviousness requirements in industries that innovate more rapidly. This corresponds to a smaller region of nonobviousness requirements where reductions in the standard encourages additional R&D.

Consider the following example. There are two industries that are identical in every respect except their R&D input prices, where $p_1 < p_2$. The analysis in section 5 shows that firms in the first industry conduct more R&D than firms in the second. Consequently, the first industry innovates more rapidly than the second. The arrival rate of patentable discoveries will also be higher. From Proposition 6, we know the R&D maximizing standard of nonobviousness is not the same for both industries. In fact, $s_1^* > s_2^*$, implying that the static region of the first industry is smaller than the static region of the second. Depending on the initial nonobviousness requirement, reductions in this requirement are more likely to *reduce* R&D activity in the first industry than in the second. In addition, the R&D maximizing standard of nonobviousness protects a smaller proportion of discoveries in the first industry than in the second. Consequently, the objectives of maximizing R&D activity and extending patent protection to a larger share of all discoveries are least compatible in rapidly innovating industries.

7. Conclusions

This paper describes an environment where the profitability of inventions is eroded by the introduction of new, competing technologies through time. When firms can readily duplicate each other's discoveries, the nonobviousness requirement plays an important role in determining the proportion of discoveries that do not affect the profits earned by proprietary discoveries. The nonobviousness requirement affects the value of patents by determining the average profitability of patentable discoveries and the expected duration of these profits.

We show that in such an environment, there exists a unique standard of nonobviousness that maximizes the rate of innovation in an industry. This critical standard also tells us the share of all possible values of the standard where relaxing the nonobviousness requirement will raise R&D activity. Finally, we show that a reduction in the nonobviousness requirement is less likely to raise R&D activity in industries that already innovate rapidly.

The actual change in R&D activity caused by a reduction in the nonobviousness requirement depends on industry characteristics and the initial stringency of the standard. This model cannot show that the actions taken in the 1980s will lead to less R&D activity. But it does show that such an outcome is possible and that the likelihood of such an outcome is higher for rapidly innovating industries. It also predicts that R&D activity may increase in certain industries and decline in others. Contrary to the conventional wisdom, reductions in the nonobviousness requirement are more likely to encourage innovation in industries that innovate slowly than in industries that innovate rapidly. Consequently, these changes could favor traditional industries over the high technology industries these proposals were designed to encourage. Given that policymakers were particularly concerned about high technology industries, there should be some concern about whether the new intellectual property regime has helped or, in fact, has made the problem worse.

There is ample room for development of the theoretical model. Perhaps the most fruitful extension is to reintroduce patent breadth, which allows us to incorporate the issues of patent infringement and of cross licensing.[45] By introducing a fixed cost to R&D programs, we can allow for entry and evaluate how nonobviousness requirements affect the equilibrium number of firms in an industry. By allowing for heterogeneity in the distribution over invention magnitudes, we can evaluate how nonobviousness requirements affect otherwise identical industries. Then we can analyze how a uniform patent law "favors" certain types of industries over others. The property rights constructed here can easily be introduced into models of endogenous growth. Finally, we can examine how the standard of nonobviousness affects collusive equilibria, where firms control the level of R&D competition to maximize the value of their patents.

The final assessment of the intellectual property reforms of the 1980s is an empirical question. Given that a decade has passed since most of the changes were introduced, there should now be adequate data to test whether a structural change in R&D activity has occurred. The model suggests a strong testable implication: R&D activity in industries that traditionally innovate slowly should rise by more than any increase in R&D for rapidly innovating industries.

Empirical analysis is complicated by the fact that there were so many changes, including tax changes, attributing any structural change to particular reforms may prove difficult. Another complication is ambiguity in the measurement of R&D intensity. Any attempt to distinguish between industries on the basis of R&D intensity must take into account input prices and productivity. Innovation can be measured by outputs, such as patents. But this is also problematic, as the definition of patents has changed and certain industries can obtain patents more easily than others.

---

[45] O'Donoghue (98) does this in an environment where invention magnitudes are chosen with certainty. It would be interesting to incorporate his definition of leading and lagging breadth in the context of stochastic invention magnitudes and examine how the welfare results change.

BIBLIOGRAPHY

Aghion, Philippe and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*. Vol. 60: 323-51.

Banner, Donald W. 1986. "The Creation of the Federal Circuit Court of Appeals and the Resulting Revitalization of the Patent System." *Albany Law Review*. Vol. 50: 585-91.

Cadot, Olivier and Lippman, Steven A. 1995. "Barriers to Imitation and the Incentive to Innovate." mimeo, INSEAD.

Coolley, Ronald B. 1994. "The Status of Obviousness and How to Assert It as a Defense." *Journal of the Patent and Trademarks Office Society*. Vol. 76: 625-44.

_____. 1989. "What the Federal Circuit Has Done and How Often: Statistical Study of the CAFC Patent Decisions - 1982 to 1988." *Journal of the Patent and Trademarks Office Society*. Vol. 71: 385-98.

Dasgupta, Partha and Joseph Stiglitz. 1980. "Uncertainty, Industrial Structure, and the Speed of R&D." *The Bell Journal of Economics*. Vol. 11: 1-28.

Desmond, Robert. 1993. "Nothing Seems Obvious to the Court of Appeals for the Federal Circuit: The Federal Circuit, Unchecked by the Supreme Court, Transforms the Standard of Obviousness Under the Patent Law." *Loyola of Los Angeles Law Review*. Vol. 26: 455-90.

Dunner, Donald R. 1985. "Introduction." *AIPLA Q. J.* Vol. 13: 185-94.

Fudenberg, Drew and Jean Tirole. 1991. *Game Theory*. Cambridge, MA: The MIT Press.

Gilbert, Robert and Carl Shapiro. 1990. "Optimal Patent Length and Breadth." *The RAND Journal of Economics*. Vol. 21:106-12.

Green, Jerry and Suzanne Scotchmer. 1995. "On the Division of Profit in Sequential Innovation." *The RAND Journal of Economics*. Vol. 26: 20-33.

_____. 1992. "Antitrust Policy, The Breadth of Patent Protection and the Incentive to Develop New Products." mimeo, University of California, Berkeley.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: The MIT Press.

Kastriner, Lawrence G. 1991. "The Rival of Confidence in the Patent System." *Journal of the Patent and Trademarks Office Society*. Vol. 73: 5-23.

Klemperer, Paul. 1990. "How Broad Should the Scope of Patent Protection Be?" *The RAND Journal of Economics*. Vol. 21: 113-30.

Lach, Saul and Rafael Rob. 1996. "R&D, Investment and Industry Dynamics." *Journal of Economics and Management Strategy*. Vol. 5: 217-49.

Lee, Tom and Louis L. Wilde. 1980. "Market Structure and Innovation: A Reformulation." *Quarterly Journal of Economics*. Vol. 94: 429-36.

Loury, Glenn C. 1979. "Market Structure and Innovation." *Quarterly Journal of Economics*. Vol. 93: 395-410.

Mansfield, Edwin, Mark Schwartz, and Samuel Wagner. 1981 . "Imitation Costs and Patents: An Empirical Study." *Economic Journal*. Vol. 91: 907-18.

Merges, Robert P. 1992. "Uncertainty and the Standard of Patentability." *High Technology Law Journal*. Vol. 7: 1-70.

_____. 1988. "Commercial Success and Patent Standards: Economic Perspectives on Innovation." *California Law Review*. Vol. 76: 805-76.

Nordhaus, W. 1969. *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change*. Cambridge, MA: MIT Press.

_____. 1972. "The Optimum Life of a Patent: A Reply." *American Economic Review*. Vol. 62: 428-31.

O'Donoghue, Ted. 1998. "A Patentability Requirement for Sequential Innovation." *The RAND Journal of Economics*. Vol. 29: 654-679.

Raskind, Leo J. 1985. "Reverse Engineering, Unfair Competition, and Fair Use." *Minnesota Law Review*. Vol. 70: 385-415.

_____ and Richard H. Stern. 1985. “Introduction." *Minnesota Law Review*. Vol. 70: 263-9.

Reinganum, Jennifer F. 1985. "Innovation and Industry Evolution." *Quarterly Journal of Economics*. Vol. 100: 81-99.

_____. 1989. "The Timing of Innovation: Research, Development, and Diffusion." in R. Schmalensee and R. D. Willig, eds. *The Handbook of Industrial Organization*, Vol. I. Amsterdam: Elsevier Science Publishers.

Risberg, Robert L., Jr. 1990. "Five Years Without Infringement Litigation Under the Semiconductor Chip Protection Act: Unmasking the Spectre of Chip Piracy in an Era of Diverse and Incompatible Process Technologies." *Wisconsin Law Review*. Vol. 24: 241-277.

Scherer, F. 1972. "Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation." *American Economic Review*. Vol. 62: 422-427.

Scotchmer, Suzanne. 1996. "Protecting Early Innovators: Should Second Generation Products Be Patentable?" *The RAND Journal of Economics*. Vol. 27: 322-31.

_____. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives*. Vol. 5: 29-41.

Scotchmer, Suzanne and Jerry Green. 1990. "Novelty and Disclosure in Patent Law." *The RAND Journal of Economics*. Vol. 21: 131-46.

Sobel, Gerald. 1988. "The Court of Appeals for the Federal Circuit: A Fifth Anniversary Look at Its Impact on Patent Law and Litigation." *The American University Law Review*. Vol. 37: 1087-1139.

Stern, Richard H. 1985. "Determining Liability for Infringement of Mask Work Rights Under the Semiconductor Chip Protection Act." *Minnesota Law Review*. Vol. 70: 272-383.

Szczepanski, Steven Z. 1987. "Licensing or Settlement: Deferring the Fight to Another Day." *AIPLA Q.J*. Vol. 15: 299-323.

APPENDIX

Proposition 1 - Suppose the R&D technology satisfies the following assumptions:

(i) $h \in [0, \bar{h}]$ s.t. $\bar{h}$, $C(\bar{h}) < \infty$;

(ii) $C'(h) > 0$, $C''(h) > 0$, $\forall\, h > 0$;

(iii) $\lim_{h \to 0} C(h)/h = \lim_{h \to 0} C'(h) = 0$;

(iv) $\lim_{\bar{h} \to \infty} C'(\bar{h}) = \infty$;

(v) $C'(h)h - C(h) \geq 0$;

(vi) $C''(h)h - C'(h) \geq 0$; and

(vii) $\lambda(n+1) \geq 1$.

Then there exists a unique symmetric stationary equilibrium of the game.

Proof: The proof is constructed through the lemmas that follow.

Lemma 1 - Suppose $V^w_{k+1} \in (0,\infty)$ and $V^w_{k+1} - V^l_{k+1} \geq 0$. Then challengers actively compete in the stage games.

Proof: We need to show that for each level of rivalry $a^i_k \in [0,\infty]$ there is at least one level of effort $h \in (0,\infty]$ s.t. $V^i_k(h, a^i_k) \geq V^i_k(0, a^i_k)$. The inequality is satisfied when:

$$\frac{\lambda\left[r \cdot V^w_{k+1} + \lambda a^i_k \cdot (V^w_{k+1} - V^l_{k+1})\right]}{\lambda a^i_k + r} \geq \frac{pC(h)}{h},$$

i.e., the minimum average cost of R&D is not too high. The third assumption assures us that the condition is satisfied for at least one strictly positive level of effort.■

Lemma 2 - If $V^w_{k+1} \in (0,\infty)$ and $V^w_{k+1} - V^l_{k+1} \geq 0$, there exists an interior equilibrium of the stage game.

Proof: The structure of the stage games is similar to the model found in Reinganum (85). The proof of existence is a modification of the proof for her model. We assume that a firm's choice of R&D effort affects the likelihood of winning and the expected length of the patent race, but does not affect the expected values of playing optimally in future races. The derivative of the firm's objective function, $\partial V^i_k / \partial h^i_k$, is

$$\frac{r\cdot\left[\lambda V_{k+1}^{w}-pC'(h_k^i)\right]+\lambda a_k^i\cdot\left[\lambda[V_{k+1}^{w}-V_{k+1}^{l}]-pC'(h_k^i)\right]+\lambda p\left[C(h_k^i)-C'(h_k^i)\cdot h_k^i\right]}{\left[\lambda(h_k^i+a_k^i)+r\right]^2}. \qquad \text{[A1]}$$

The numerator of [A1], which we denote $\varphi^i(h_k^i,a_k^i)$, determines the sign of $\partial V_k^i(h_k^i,a_k^i)/\partial h_k^i$. But $\varphi^i(h_k^i,a_k^i)$ is strictly decreasing in R&D intensity:

$$\frac{\partial\varphi^i(h_k^i,a_k^i)}{\partial h_k^i}=-pC''(h_k^i)\cdot\left[\lambda(h_k^i+a_k^i)+r\right]<0. \qquad \text{[A2]}$$

If the saturation point of R&D ($\bar{h}$) is sufficiently large, there will be a finite level of R&D effort, denoted $\hat{h}_k^i<\bar{h}$, where $\varphi^i(\hat{h}_k^i,a_k^i)\leq 0$. For any level of effort $h_k^i>\hat{h}_k^i$, the firm's objective function is declining. Continuity of the objective function implies that $V_k^i(h_k^i,a_k^i)$ is maximized by the level of R&D effort, contained in the interval $(0,\hat{h}_k^i]$, where $\varphi^i(h_k^i,a_k^i)=0$. Let $h_k^i(a_k^i)$ denote the firm's best response to the level of rivalry it encounters. The strict monotonicity of $\varphi^i(h_k^i,a_k^i)$ implies that this best response is unique.

Firms never choose R&D intensities greater than $\hat{h}_k^i$, so we can restrict the strategy space to a convex, compact, nonempty subset of $\mathbb{R}^n$, denoted $X\equiv\Pi_{i=1}^n[0,\hat{h}_k^i]$. The vector $[h_k^1(a_k^1),\ h_k^2(a_k^2),\ ...,\ h_k^n(a_k^n)]$ maps X into itself continuously. Existence of an equilibrium then follows from Brouwer's fixed point theorem.■

Lemma 3 - If $\lambda[V_{k+1}^{w}-V_{k+1}^{l}]-p[C'(h_k)+h_k\cdot C''(h_k)]<0,$ there exists a unique, symmetric equilibrium of the stage game.

Proof: Existence of a symmetric equilibrium follows from the objective functions and first order condition of firms, which varies only by the level of rivalry encountered. In the symmetric equilibrium, $\varphi^i(h_k^i,a_k^i)$ becomes $\varphi^i(h_k,(n-1)\cdot h_k)$. The corresponding first order condition is

$$r\cdot\left[\lambda V_{k+1}^{w}-pC'(h_k)\right]+\lambda(n-1)h_k\cdot\left[\lambda[V_{k+1}^{w}-V_{k+1}^{l}]-pC'(h_k)\right]+\lambda p\left[C(h_k)-C'(h_k)\cdot h_k\right]=0.$$

The first and third terms are strictly decreasing in R&D effort. If the second term is also strictly decreasing, then only one level of R&D intensity satisfies the equality. Hence we require that

$$\lambda\left[V_{k+1}^{w}-V_{k+1}^{l}\right]-p\left[C'(h_k)+h_k\cdot C''(h_k)\right]<0.\ ■$$

The symmetric equilibrium R&D intensity of the stage game with continuation values V_{k+1}^{w} and V_{k+1}^{l} is denoted $h_k(V_{k+1}^{w},V_{k+1}^{l})$.

Lemma 4 - The game is continuous at infinity.

Proof: It is sufficient to show that total firm payoffs are a discounted sum of per period payoffs and that these per period payoffs are uniformly bounded [see Fudenberg and Tirole (91), p. 110]. The per period payoff to firms is the present value of flow profits for the incumbent and the present value of R&D expenditures for challengers. The maximum per period return for an incumbent is $\bar{u}/r$. Per period returns for challengers are contained in the interval $[-C(\bar{h})/(r + \bar{h}), 0]$.■

Lemma 5 - Lemmas 1 - 4 imply the existence of a stationary symmetric equilibrium of the game.

Proof: We return to the first order condition of the stage game, but assume that the continuation values associated with winning and losing the current race do not vary across races. Rearranging terms, we have:

$$pC'(h_k)\cdot\left[r+\lambda n h_k\right]=\lambda\left[r\cdot V^{w}+pC(h_k)+\lambda(n-1)h_k\cdot[V^{w}-V^{l}]\right]. \qquad \text{[A3]}$$

If firms take the continuation values as given, and these values are constant across races, it is a best response for each firm to choose the same R&D intensity $h_k = h(V^w, V^l)$ in each race. Lemma 3 establishes the existence of such a best response for a given specification of the continuation values. We continue to assume that the best response is unique and will later verify that the necessary condition is satisfied. In section 5b of the paper, we computed the corresponding continuation values:

$$V^{I}(h)=\frac{\left[r+\theta\lambda h\right]\cdot u^{e}-\theta\lambda nh\cdot pC(h)}{r\cdot[r+\theta\lambda(n+1)h]}, \qquad \text{[A4]} \qquad\qquad V^{C}(h)=\frac{\theta\lambda h\cdot u^{e}-\left[r+\theta\lambda nh\right]\cdot pC(h)}{r\cdot[r+\theta\lambda(n+1)h]}. \qquad \text{[A5]}$$

An equilibrium of the game is described by the best response $h(V^w, V^l)$ where $V^l = V^C(h)$ and $V^w = \theta V^I(h) + (1-\theta)V^C(h)$. Substituting [A4] and [A5] into [A3] yields the equilibrium condition

$$pC'(h)=\theta\cdot[V^{I}-V^{C}]=\theta\cdot\left(\frac{u^{e}+pC(h)}{r+\theta\lambda(n+1)h}\right). \qquad \text{[A6]}$$

We use σ to denote the equilibrium R&D intensity that satisfies [A6]. Note that [A6] is consistent with the condition required in lemma 3 for the uniqueness of symmetric equilibrium strategies of the stage games.

During each race, for every firm the R&D intensity σ is the unique best response to the continuation values $V^I(\sigma)$ and $V^C(\sigma)$. Then the strategy of playing σ in every race cannot be improved upon by choosing a different R&D intensity in one race and playing σ in all the others. Finally, if playing σ in every race cannot be improved upon by a deviation in one stage, and the game is continuous at infinity, choosing the R&D intensity σ in each race is a subgame perfect equilibrium of the game [see Fudenberg and Tirole (91), p. 110].■

Lemma 6 - The symmetric stationary equilibrium is unique.

Proof: It is sufficient to show that there is only one possible intersection of the curves described by $pC'(h)$ and $\theta[V^I(h) - V^C(h)]$. The difference in the slopes of the two sides is

$$p\left(\frac{c''(h)[r+\theta\lambda(n+1)h] + C'(h)\theta[\lambda(n+1)-1]}{r+\theta\lambda(n+1)h}\right) \equiv \frac{\partial\Omega(h)}{\partial h}, \qquad \text{[A7]}$$

where $\Omega(h) = pC'(h) - \theta[V^I(h) - V^C(h)]$. At $h = 0$, $pC'(h)$ is zero while $\theta[V^I(h) - V^C(h)]$ is u^e/r. The existence of an equilibrium implies that [A7] is nonnegative at the first intersection of the curves. If there are multiple intersections, for at least one intersection $\theta[V^I(h) - V^C(h)]$ must be increasing faster than marginal cost. In that equilibrium, the numerator of [A7] is negative. A necessary, but not sufficient, condition for such a possibility is that the R&D productivity coefficient is very small [$\lambda < 1/(n+1)$]. But this contradicts assumption (vii).■

Proposition 2 - The rate of innovation in an industry is increasing in the number of firms and the productivity of R&D, but decreasing in the discount rate and the price of R&D inputs.

Proof: These results depend on the comparative static properties of the equilibrium:

i. Increasing the number of firms:

$$\frac{\partial\sigma}{\partial n} = \frac{-C'(\sigma)\theta\lambda\sigma}{C''(\sigma)[r+\theta\lambda(n+1)\sigma] + C'(\sigma)\theta[\lambda(n+1)-1]} < 0;$$

which implies that per firm R&D effort declines. However, the industrywide arrival rate of discoveries depends on $n\sigma$. Increasing the number of firms changes the rate of innovation in an industry by $\sigma + n\cdot\partial\sigma/\partial n$, which is

$$\frac{\partial n\sigma}{\partial n} = \sigma\cdot\left(\frac{C''(\sigma)[r+\theta\lambda(n+1)\sigma] + C'(\sigma)\theta(\lambda-1)}{C''(\sigma)[r+\theta\lambda(n+1)\sigma] + C'(\sigma)\theta[\lambda(n+1)-1]}\right).$$

This derivative is nonnegative if $C''(\sigma)\sigma - C'(\sigma) \geq 0$ and $\lambda \geq 1/(n+1)$.

ii. Increasing the productivity of R&D:

$$\frac{\partial\sigma}{\partial\lambda} = \frac{-C'(\sigma)\theta(n+1)\sigma}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]} < 0;$$

which implies that per firm R&D effort declines. But the industrywide rate of discovery depends on $\lambda\sigma$. Raising the productivity parameter causes the rate of discovery to change by $\sigma+\lambda\cdot\partial\sigma/\partial\lambda$, which is

$$\frac{\partial\lambda\sigma}{\partial\lambda} = \sigma\left(\frac{rC''(\sigma)+\theta[C''(\sigma)\lambda(n+1)\sigma-C'(\sigma)]}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]}\right).$$

This derivative is non-negative if $C''(\sigma)\sigma$ - $C'(\sigma)\geq 0$ and $\lambda\geq 1/(n+1)$.

iii. Increasing the discount rate:

$$\frac{\partial\sigma}{\partial r} = \frac{-C'(\sigma)}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]} < 0.$$

iv. Increasing the relative price of R&D inputs:

$$\frac{\partial\sigma}{\partial p} = \frac{-1}{p}\cdot\left(\frac{rC'(\sigma)+\theta[C'(\sigma)\lambda(n+1)\sigma-C(\sigma)]}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]}\right).$$

Higher input prices cause R&D activity to decline if $C'(\sigma)\sigma$ - $C(\sigma)\geq 0$ and $\lambda\geq 1/(n+1)$. To see if actual expenditures rise or decline, we examine the derivative $\partial p\sigma/\partial p$:

$$\frac{\partial p\sigma}{\partial p} = \frac{[C''(\sigma)\sigma-C'(\sigma)][r+\theta\lambda(n+1)\sigma]+\sigma C'(\sigma)\theta[\lambda(n+1)-1]+\theta C(\sigma)}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]}.$$

This derivative is nonnegative if $C''(\sigma)\sigma$ - $C'(\sigma)\geq 0$ and $\lambda\geq 1/(n+1)$.■

Proposition 3 - The derivation of $\Psi(s)$.

Proof: We must check the following derivative:

$$\frac{-\partial\Omega(\sigma)}{\partial s} = [V^{I}(\sigma)-V^{C}(\sigma)]\cdot\frac{\partial\theta}{\partial s}+\theta\cdot\left(\frac{\partial V^{I}(\sigma)}{\partial s}-\frac{\partial V^{I}(\sigma)}{\partial s}\right). \qquad \text{[A8]}$$

Recall that θ = 1- F(s), which implies that $\partial\theta/\partial s = -f(s)$. The other derivatives are

$$\frac{\partial V^I(\sigma)}{\partial s} = \frac{f(s)}{\theta}\cdot\left\{\left(\frac{r+\theta\lambda\sigma}{r+\theta\lambda(n+1)\sigma}\right)\cdot\frac{[u^e(s)-s]}{r}+\left(\frac{\theta\lambda n\sigma}{r+\theta\lambda(n+1)\sigma}\right)\cdot[V^I(\sigma)-V^C(\sigma)]\right\},$$

and

$$\frac{\partial V^C(\sigma)}{\partial s} = \frac{f(s)}{\theta}\cdot\left(\frac{\theta\lambda\sigma}{r+\theta\lambda(n+1)\sigma}\right)\cdot\left\{\frac{[u^e(s)-s]}{r}-[V^I(\sigma)-V^C(\sigma)]\right\}.$$

Taking the difference of these equations yields:

$$\frac{f(s)}{\theta}\cdot\left\{\left(\frac{r}{r+\theta\lambda(n+1)\sigma}\right)\cdot\frac{[u^e(s)-s]}{r}+\left(\frac{\theta\lambda(n+1)\sigma}{r+\theta\lambda(n+1)\sigma}\right)\cdot[V^I(\sigma)-V^C(\sigma)]\right\}. \qquad \text{[A9]}$$

Plugging [A9] into [A8] and rearranging terms, the comparative static calculation becomes

$$\frac{\partial\sigma}{\partial s} = \frac{f(s)\Psi(s)}{p\big(C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]\big)}; \qquad \text{[A10]}$$

where

$$\Psi(s) \equiv \left(\frac{\theta\lambda(n+1)\sigma}{r+\theta\lambda(n+1)\sigma}\right)\cdot[u^e(s)-s] - \left(\frac{r}{r+\theta\lambda(n+1)\sigma}\right)\cdot[s+pC(\sigma)]. \ \blacksquare$$

Proposition 5 - There is a unique standard of nonobviousness, denoted s^*, such that in the interval $[0, s^*)$, R&D activity is strictly increasing in the standard of nonobviousness.

Proof: We begin by examining the derivative $\partial\Psi(s)/\partial s$:

$$\left(\frac{\theta\lambda(n+1)\sigma}{r+\theta\lambda(n+1)\sigma}\right)\cdot\frac{f(s)}{\theta}\cdot\Psi(s) + \left(\frac{r}{r+\theta\lambda(n+1)\sigma}\right)pC'(\sigma)[\lambda(n+1)-1]\cdot\frac{\partial\sigma}{\partial s} - 1. \qquad \text{[A11]}$$

Substituting the expression for $\partial\sigma/\partial s$, [A11] becomes

$$\frac{\partial\Psi(s)}{\partial s}=\Psi(s)\cdot\frac{f(s)}{\theta}\cdot\left(\frac{C''(\sigma)\theta\lambda(n+1)\sigma+C'(\sigma)\theta[\lambda(n+1)-1]}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]}\right)-1.$$

If there exists an extremum of σ(*s*) for some $s*\in[0,\bar{u}]$, then Ψ(s*) = 0 and $\partial\Psi(s)/\partial s|_{s*}=-1$. Then, for all values *s* > *s**, Ψ(s) < 0, implying that ∂σ/∂*s* < 0. Thus there can be at most one extremum of σ(*s*). Next, we check the values of Ψ(s) as s → 0 and s → $\bar{u}$. Then we see that $\lim_{s\to\bar{u}}\Psi(s)\leq-\bar{u}$. The other limit is

$$\lim_{s\to 0}\Psi(s)=\frac{\lambda(n+1)\sigma(0)\cdot u^{e}(0)-r\cdot pC(\sigma(0))}{r+\lambda(n+1)\sigma(0)}. \qquad [A12]$$

To verify the sign of this expression, we evaluate the condition that ensures firms actively compete in the patent races when *s* = 0.

Lemma 7 - Lemma 1 implies that $pC(\sigma(0))\leq u^{e}(0)\cdot\lambda\sigma(0)/[r+\lambda n\sigma(0)]$.

Proof: When s = 0, so θ = 1, the constraint derived in lemma 1 is

$$pC'(h_k)\cdot\left[r+\lambda nh_k\right]=\lambda\left[r\cdot V^{w}+pC(h_k)+\lambda(n-1)h_k\cdot[V^{w}-V^{l}]\right].$$

Because all innovations are patentable, $V^w = V^I(\sigma(0))$ and $V^l = V^C(\sigma(0))$. Substituting these continuation values into the constraint yields

$$\frac{[r+\lambda n\sigma(0)]u^{e}(0)-\lambda\sigma(0)pC(\sigma(0))}{r+\lambda(n+1)\sigma(0)}\geq\frac{pC(\sigma(0))[r+\lambda(n-1)\sigma(0)]}{\sigma(0)}.$$

Finally solving for *pC*(σ(0)) and simplifying terms, we arrive at

$$\left(\frac{\lambda\sigma(0)}{r+\lambda n\sigma(0)}\right)\cdot u^{e}(0)\geq pC(\sigma(0)).\ ■$$

By substituting $u^{e}(0)\cdot\lambda\sigma(0)/[r+\lambda n\sigma(0)]$ for *pC*(σ(0)) in [A12], we can show that

$$\lim_{s\to 0}\Psi(s)\geq\left(\frac{\lambda n\sigma(0)}{r+\lambda n\sigma(0)}\right)\cdot u^{e}(0)>0. \qquad [A13]$$

The continuity of Ψ(s) and the signs of Ψ(0) and Ψ($\bar{u}$) imply the existence of at least one standard $s*\in[0,\bar{u}]$, where Ψ(s*) = 0. ■

Proposition 6 - The critical standard of nonobviousness, s*, is increasing in the equilibrium rate of innovation.

Proof: The critical standard of nonobviousness is defined by the equation $\Psi(s) = 0$. To examine how s^* varies with the rate of innovation, we compute comparative static derivatives with respect to the parameters (identified earlier) related to an industry's equilibrium rate of innovation:

$$\frac{\partial s*}{\partial z} = -\frac{\partial \Psi(s)}{\partial z} \Big/ \frac{\partial \Psi(s)}{\partial s} = \frac{\partial \Psi(s)}{\partial z}.$$

i. Increasing the number of firms:

$$\frac{\partial \Psi(s*)}{\partial n} = \left(\frac{r \cdot pC'(\sigma)}{r + \theta\lambda(n+1)\sigma} \right) \cdot \left(\lambda\sigma + [\lambda(n+1) - 1] \frac{\partial \sigma}{\partial n} \right).$$

The sign is not immediately obvious, because per firm R&D intensity decreases as the number of firms rises. Substituting for $\partial\sigma/\partial n$ implies that

$$\frac{\partial \Psi(s)}{\partial n} = \frac{\lambda\sigma C''(\sigma) \cdot rpC'(\sigma)}{C''(\sigma)[r + \theta\lambda(n+1)\sigma] + C'(\sigma)\theta[\lambda(n+1) - 1]} > 0 \;\Rightarrow \frac{\partial s*}{\partial n} > 0.$$

ii. Increasing the productivity of R&D:

$$\frac{\partial \Psi(s*)}{\partial \lambda} = \left(\frac{r \cdot pC'(\sigma)}{r + \theta\lambda(n+1)\sigma} \right) \cdot \left((n+1)\sigma + [\lambda(n+1) - 1] \frac{\partial \sigma}{\partial \lambda} \right).$$

Recall that equilibrium R&D intensity is decreasing in its productivity, but the rate of innovation is increasing in R&D productivity (because σ falls less than λ rises). Substituting $\partial\sigma/\partial\lambda$ into the preceding expression yields

$$\frac{\partial \Psi(s)}{\partial \lambda} = \frac{(n+1)\sigma C''(\sigma) \cdot rpC'(\sigma)}{C''(\sigma)[r + \theta\lambda(n+1)\sigma] + C'(\sigma)\theta[\lambda(n+1) - 1]} > 0 \;\Rightarrow \frac{\partial s*}{\partial \lambda} > 0.$$

iii. Increasing the discount rate:

$$\frac{\partial \Psi(s*)}{\partial r} = -\left(\frac{pC'(\sigma)}{r + \theta\lambda(n+1)\sigma} \right) \left(\lambda(n+1)\sigma - r[\lambda(n+1) - 1] \frac{\partial \sigma}{\partial r} \right).$$

Higher discount rates are associated with lower equilibrium R&D intensity. Substituting $\partial\sigma/\partial r$, we find that

$$\frac{\partial\Psi(s)}{\partial r}=-pC'(\sigma)\cdot\left(\frac{C''(\sigma)\lambda(n+1)\sigma+C'(\sigma)[\lambda(n+1)-1]}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]}\right)<0\ \Rightarrow\frac{\partial s*}{\partial r}<0.$$

iv. Increasing the relative price of R&D inputs:

$$\frac{\partial\Psi(s*)}{\partial p}=-\left(\frac{r}{r+\theta\lambda(n+1)\sigma}\right)\left(C(\sigma)-pC'(\sigma)[\lambda(n+1)-1]\cdot\frac{\partial\sigma}{\partial p}\right).$$

But higher input prices reduce equilibrium R&D intensity. Substituting $\partial\sigma/\partial p$, we find that

$$\frac{\partial\Psi(s)}{\partial p}=-r\cdot\left(\frac{C''(\sigma)C(\sigma)+C'(\sigma)^2[\lambda(n+1)-1]}{C''(\sigma)[r+\theta\lambda(n+1)\sigma]+C'(\sigma)\theta[\lambda(n+1)-1]}\right)<0\ \Rightarrow\frac{\partial s*}{\partial p}<0.\blacksquare$$