WORKING PAPERS

RESEARCH DEPARTMENT

WORKING PAPER NO. 01-13/R
PATENTABILITY, INDUSTRY STRUCTURE, AND INNOVATION

Robert M. Hunt
Federal Reserve Bank of Philadelphia

First draft: August 2001
Revised: August 2004

FEDERAL RESERVE BANK OF PHILADELPHIA

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

Working Paper No. 01-13/R

PATENTABILITY, INDUSTRY STRUCTURE, AND INNOVATION

Robert M. Hunt*
Federal Reserve Bank of Philadelphia

First Draft: August 2001
Revised: August 2004

* Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106. Phone: 215-574-3806. Email: bob.hunt@phil.frb.org

The author wishes to especially thank George Mailath for his comments and insights on many previous drafts of this paper. Thanks also to Stephen Coate, Dennis Yao, Leonard Nakamura, Nikolaos Vettas, Alberto Trejos, Chris Chalmers, Juuso Valimaki, Kamal Saggi, and Rohit Verma, and two anonymous referees. Any remaining errors are, of course, my own.

PATENTABILITY,
INDUSTRY STRUCTURE,
AND INNOVATION

Robert M. Hunt*
Federal Reserve Bank of Philadelphia

August 2004

Abstract

This paper presents a model of sequential innovation in which industry structure is endogenous and a standard of patentability determines the proportion of all inventions that qualify for protection (in U.S. patent law this standard is called *nonobviousness*; in Europe it is called the *inventive step*). The rate of innovation initially rises as this standard is raised from very low levels, but eventually falls as the standard is raised to very high levels. Hence, there is a unique patentability standard that maximizes the rate of innovation. Surprisingly, this critical standard is more stringent for industries disposed to innovate rapidly. The model suggests a number of important implications for patent policy.

Keywords: Patents, intellectual property, nonobviousness, inventive step
JEL Codes: L1, O31, O34

* Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106. Phone: 215-574-3806. Email: bob.hunt@phil.frb.org

The author wishes to especially thank George Mailath for his comments and insights on many previous drafts of this paper. Thanks also to Stephen Coate, Dennis Yao, Leonard Nakamura, Nikolaos Vettas, Alberto Trejos, Chris Chalmers, Juuso Valimaki, Kamal Saggi, and Rohit Verma, and two anonymous referees. Any remaining errors are, of course, my own.

1. Introduction

Recently, economists have investigated, in the context of cumulative innovation, the relationship between the availability of patent protection and the rate of innovation (Bessen and Maskin 2002, Hunt 1999a, O'Donoghue 1998, and Scotchmer 1996). The general conclusion is that an industry's rate of innovation is maximized by protecting some inventions, but not others.

This paper presents a model in which industry structure (the number of firms engaged in R&D) depends in part on the share of all discoveries that qualify for protection, that is, by the stringency of the criteria used to examine applications for a patent. In the model, the number of firms actively engaged in R&D is the primary determinant of an industry's rate of progress. This in turn depends on the fixed cost of establishing a research facility, the productivity of R&D, and the resulting profits generated in the output market. Patentability standards affect expected profits because they determine the likelihood that a firm's invention will lead to a competitive advantage and the speed with which that advantage will be eroded.

When we speak of a standard of patentability in this paper, we focus on American patent law's requirement of *nonobviousness*, or what is called the *inventive step* in Europe. To qualify for a patent it is not sufficient for an invention to be new; it must also represent a sufficiently large advance from the prior art. One can think intuitively of the nonobviousness requirement as specifying the minimal advance—the 'height' of the inventive step—necessary to qualify for protection.

In the model, industry structure is characterized by a single firm in the output market that is eventually replaced by a firm that develops a patentable innovation. We show that the arrival rate of these innovations is a non-monotonic function of the stringency of the patentability standard applied to inventions. There is a unique critical value of this standard where the rate of innovation in an industry is maximized. This is accomplished by maximizing the number of firms that choose to engage in R&D. The critical patentability standard depends on exogenous parameters that influence an industry's propensity to innovate. The critical patentability standard is more stringent in industries otherwise pre-disposed to innovate rapidly and less stringent for industries predisposed to innovate more slowly. In other words, in order to maximize the rate of innovation in "hi-tech" industries, one should require a relatively tall inventive step and not a shorter one.

The remainder of the paper is organized as follows. Section 2 introduces the model and compares it to the existing literature. Section 3 presents the equilibrium and describes its properties. Section 4 describes the relationship between the inventive step and the rate of innovation and derives the R&D maximizing standard as a function of the exogenous parameters that determine an industry's propensity to innovate. It also presents the first and second best solutions to the social planner's problem. Section 5 examines the policy implications of these results. The Appendix contains the proofs of all the propositions.

2. The Model

2.1 An Infinite Sequence of Stochastic Patent Races

Time is continuous and the horizon is infinite. Let $r > 0$ denote the discount rate. Discoveries occur at different points in time. Time is divided into the intervals between discoveries. Each interval is a patent race. The duration of patent races varies because there is randomness in the process that generates discoveries.

There are $n_t + 1$ firms in the industry, where $n_t \geq 0$ is determined by a free entry constraint that depends on the fixed cost ($k > 0$) of setting up an R&D lab. This cost is sunk when the firm enters its first patent race; new fixed investments are not required thereafter.

Firms are indexed by the superscript i. At the beginning of a race, firms simultaneously choose their R&D intensity, denoted $h^i \in [0, \bar{h}]$, where $\bar{h}$ is a very large, but finite, point of saturation. Firms maintain their research intensity until a discovery occurs and the current race ends. The flow cost of conducting R&D, denoted $C(h^i)$, is strictly increasing and twice continuously differentiable in R&D intensity.

All firms share the same R&D technology. A firm's discoveries arrive through time according to a Poisson process, where the arrival rate is an increasing function of its R&D intensity. Denote the arrival rate of ideas for firm i in patent race q as λh_q^i, where λ is an industry-specific productivity parameter. The probability that firm i discovers an invention before date t in the patent race q is $1 - e^{-\lambda \cdot h_q^i \cdot t}$. The firm faces a constant rival hazard rate $\lambda a_q^i \equiv \lambda \sum_{j \neq i} h_q^j$. The probability that firm i wins patent race q is $h_q^i / [h_q^i + a_q^i]$, the ratio of firm i's hazard rate to the hazard rate for the entire industry.

2.2 A Passive Incumbent

A firm that owns a patented invention will be called an *incumbent*. The other firms will be called *challengers*. The model contains an additional assumption about the nature of technological competition: A firm that makes a patentable discovery does not compete in the subsequent patent race.

This ad hoc restriction considerably simplifies the model and subsequent analysis, but it does not affect the qualitative properties of a model of patent races where the only difference between the incumbent and other firms is the rents it earns. In models of this type, being successful in a given patent race does not convey a natural advantage over rivals in subsequent races. It can be shown that the incumbent will race less aggressively than other firms, because it takes into account the fact that its R&D may replace profits it already earns (Reinganum 1985). In other models (Grossman and Helpman 1991), incumbents do not race at all.[1]

2.3 The Nature of Inventions and a System of Property Rights

A discovery is an improvement in product quality. The extent of an improvement is denoted $u_q \in [0, \overline{u}]$,where $\overline{u} < \infty$.[2] 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.

Once a discovery has been made, it can be reverse-engineered at zero cost by all other firms. If a patent is granted, the inventor receives an exclusive right to produce and sell that invention. The statutory life of the patent is infinite. Not all inventions will be protected, however. Let $s \in [0, \overline{u}]$ denote the minimum extent of improvement for which the patent office is willing to grant a patent. In the model, this is the inventive step or standard of nonobviousness. An invention whose extent is less than s is not protected and becomes available to all firms.[3] In other words, it is added to the public domain of product improvements. Let $\theta(s) = 1 - F(s)$ denote the ex-ante probability of obtaining patent protection, given the patentability

[1] In that model firms borrow to finance their R&D investments, and the arrival rate of innovations is linear in firms' investments. In that case, the incumbent is at such a disadvantage vis-à-vis its rivals that it cannot finance subsequent innovations. It should be noted that in models that contain more asymmetry between firms, the assumption made here would significantly affect the properties of the resulting equilibria.

[2] Alternatively, we can express innovations as some percent reduction in the cost of producing the final good. The analysis would yield the same results so long as we assume cost reductions are perfectly compatible, so that a cost reduction applied to different vintages of technology achieves the same percent reduction in cost.

[3] In the typology of O'Donoghue, Scotchmer, and Thisse (1998), we assume that lagging breadth is equivalent to the magnitude of any patentable invention, while there is no leading breadth.

standard s. The expected quality improvement of a patentable invention will be denoted

$\tilde{u} = \int_s^{\bar{u}} u dF(u) / [1 - F(s)]$.

Patent claims are defined as the improvement itself, so each improvement does not infringe a patent on another improvement. But when, and under what conditions, will an inventor be able to use prior generations of improvements in her product? For example, the firm might be required to license all prior improvements from their inventors. At the other extreme, an inventor could use all prior discoveries without obtaining a license. In this paper, we adopt an intermediate case: if an invention satisfies the standard of patentability, the inventor may use all prior discoveries without licensing them. 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. 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.[4]

Lach and Rob (1996) adopt an alternative approach, where firms embody new technology in vintage-specific capital goods. 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. In the model of O'Donoghue (1998), owners of patented inventions must cross license with each other if they are to produce a final good using the best available technology. To reduce complexity, O'Donoghue assumes such licenses are achieved, but at the expense of an exogenous transactions cost. In his model, a social planner would respond to a higher transactions cost by raising the standard of patentability. If cross licensing were required in the model presented here, the same intuition would apply.

[4] This definition is consistent with the "reverse engineering" defense Congress established for *mask rights*, a sui generis form of intellectual property protecting the physical layout of computer chips (Hunt 1999a).

2.4 The Output Market and Flow Profits

From the preceding section, it is clear that during patent race q, the current holder of a patent can offer a product with the best available technology, i.e., one that embodies all the quality improvements invented prior to this race. The best any competing firm may offer is a product embodying all the improvements except for this last patented invention. Let $\hat{u}_q$ denote the extent of the innovation protected during race q. Note this is not necessarily the invention that ended the previous race.

All consumers are identical and aggregate demand is normalized to one. Consumers care only about the quality of the good they are consuming. The reservation value of the final product to consumers, then, is simply the level of its quality, multiplied by p, the price of the final good relative to the R&D inputs.[5] Firms compete in prices and the cost of production is zero. Thus the equilibrium price of the final good, and the incumbent's flow profit, during the qth race is $p \cdot \hat{u}_q$.

We are interested in the flow profits earned by firms in the next (q+1) race. Several things might happen during the current race. Suppose that challenger i invents first in the qth race, but the invention is too small to qualify for protection. Because all firms can use that invention, the competitive position of firms in the output market is unchanged.[6] In that case, the leader during race q continues to earn flow profits of $p \cdot \hat{u}_q$ in the next race while all other firms earn nothing (see the last column of the table). Alternatively, the magnitude of i's invention is sufficiently large that it qualifies for protection. According to the property rights defined earlier,

---

[5] 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.

[6] For example, research performed by NACA (the predecessor to NASA) in the 1920s and 1930s led to the development of significantly more efficient engine cowlings and airfoils. These discoveries were disclosed to all and quickly adopted by aircraft manufacturers around the world. Some of these discoveries could have been patented, but the government's policy at the time was not to do so.

firm *i* can also use all previously patented inventions. In that case, during race q+1, firm *i* will earn flow profit $p \cdot \hat{u}_{q+1} = p \cdot u_q$, while the previous leader and all other firms earn nothing (see the first column of the table).

Flow Profits Earned during Patent Race *q*+1

	Innovation q was	
The firm is	Patentable	Unpatentable
The leader from race (q-1)	0	$p \cdot \hat{u}_{q+1} = p \cdot \hat{u}_q$
The winning challenger *i*	$p \cdot \hat{u}_{q+1} = p \cdot u_q$	0
All other challengers	0	0

2.5 The Existing Literature

The model builds on the extensive literature on stochastic patent races (Loury 1979, Dasgupta and Stiglitz 1980, Lee and Wilde 1980, and Reinganum 1985). The resulting equilibrium is similar to ones analyzed in certain models of endogenous growth (Aghion and Howitt 1992 and Grossman and Helpman 1991). One can interpret these models as an extreme case of the model constructed here, when all innovations satisfy the standard of patentability and every discovery eliminates the rents associated with the prior one.

Gilbert and Shapiro (1990) and Klemperer (1990) examine the optimal combination of patent length and *breadth*. Breadth is the degree to which a product or process must differ from a patented one to avoid infringement of the patent. In contrast, the 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 be nonobvious and still infringe a prior patent.

Scotchmer and Green (1990) is one of the first papers to model the effects of a patentability standard. This line of research (which includes Green and Scotchmer 1995, Scotchmer 1996, and Denicolò 2000) examines the role of patents in the context of cumulative innovation, i.e., where inventions build on each other. These papers examine, in a two-period model, how patents should be designed to achieve an optimal allocation of rents between initial and subsequent innovators.

There are a now a number of papers that evaluate intellectual property rules in dynamic models of sequential innovation. Bessen and Maskin (2002) show that an environment without any patent protection can generate more innovation than an environment with patents. The key to this result is that inventions in their model are both complementary and essential, so that firms benefit from their rivals' R&D even if they must also share rents with them.

The finding that the rate of innovation is a non-monotonic function of the extent (or availability) of patent protection is found in several papers.[7] In O'Donoghue (1998) firms choose how much to invest in R&D and a deterministic invention size. He shows that a social planner can induce more rapid innovation by specifying a minimum invention size that qualifies for patent protection. The mechanism is similar to one described here – lengthening the duration of incumbency can increase rents and consequently stimulate R&D investments.

In Horowitz and Lai (1996), firms choose how fast to race and the extent of the innovation they are targeting. They find the market leader will innovate just before its existing patent expires and that the extent of its innovation is an increasing function of the patent term. They show that the overall rate of innovation is maximized with a patent of finite duration, but that social welfare is maximized with an even shorter patent term. In their model, patent length (measured in time) plays a role comparable to nonobviousness in this model.

The primary difference between the model presented here and the previous literature is

[7] See also the papers by Cadot and Lippman (1997) and Chou and Haller (1995). In these models, the incentive to innovate is a non-monotonic function of rivals' exogenously specified capacity to imitate.

that the magnitude of innovations is stochastic and industry structure is endogenous. In this environment, the relevant policy parameter is not patent life, which is also endogenous, but the minimum invention size that qualifies for protection. The inventive step that maximizes the rate of innovation in an industry is the one that maximizes the number of firms engaged in R&D.[8]

3. Equilibrium

In this model, the leading firm is a passive recipient of rents earned on its previous patentable discovery. Eventually an innovation will occur, ending the current race and possibly the incumbent's rents. During the current race, challengers select the R&D intensity that maximizes expected current cash flow plus the expected present value of competing optimally in future races.[9] The exact magnitude of flow profits associated with a patentable discovery is not known until the discovery has actually occurred. Firms take into account the expected magnitude of patentable discoveries $(\tilde{u})$ when choosing their R&D intensity. The challengers move simultaneously, taking the number of their rivals as given.

3.1 The Stationary Symmetric Equilibrium of the Game

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. At the beginning of each race, each firm knows the play of all firms in the prior races and the outcomes of those races. 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 $\mathbb{R}$. There are likely

[8] Bernheim (1984) shows that in industries subject to sequential entry, excessively vigorous antitrust enforcement results in more concentration, not less. The underlying mechanism is similar to the one explored in this paper – if government policies reduce the likelihood of earning significant rents, only a few firms are able to amortize their cost of entry. I am grateful to an anonymous referee for pointing out this parallel.

[9] The objective function for firms in each individual patent race is specified in the Appendix.

to be many equilibria of the game, but we focus on stationary equilibria where firms choose identical strategies. In the Appendix, we prove the following:

> **Proposition 1** - Suppose the R&D cost function satisfies the following assumptions: (i) $C(h) > 0,\ C'(h) > 0\ \forall h > 0;$ (ii) $C''(h) > 0\ \forall h > \hat{h} \in [0, \bar{h});$ (iii) $\mathrm{Lim}_{h\to 0}\ C(h)/h = \mathrm{Lim}_{h\to 0}\ C'(h) = 0;$ (iv) $\mathrm{Lim}_{h\to\infty}\ C'(h) = \infty;$ and (v) $C(\bar{h}) < \infty,\ \forall \bar{h} \in [0, \infty).$ Then, there exists a unique, stationary, symmetric equilibrium of the game, characterized by the pair $(\sigma^*, n^*),$ where there are n^* challengers who choose a flow R&D intensity $\sigma^* \in (0, \bar{h}].$

The first two assumptions tell us that R&D costs rise with intensity and R&D is eventually subject to diminishing returns. Together with the third and fourth assumptions, this ensures there will be an interior equilibrium of the stage games. To ensure the existence of a Markov Perfect Equilibrium, we need only verify that per period payoffs are bounded. This is ensured by the fifth assumption and the fact that the largest invention magnitude is finite.

The R&D technology, the distribution of invention magnitudes, and the relationship between patented technology and expected profits do not vary across races.[10] If all challengers choose the same R&D intensity in all patent races, the probabilities of winning and losing, the expected length of races, and the continuation values associated with being the incumbent or a challenger will be the same in each race.[11]

Let V^I and V^C denote the present value of a firm that is currently the incumbent and a

---

[10] A more general model would allow for exhaustion of technological opportunities or spillovers from advances in other fields. The resulting dynamics would be both complicated and interesting.

[11] Note that the environment is not completely stationary until after the second patent race: In the first race there is no incumbent; in the second race, the free entry condition implies that an additional challenger will enter. But this does not affect the qualitative properties of the equilibrium analyzed here.

challenger, respectively. In equilibrium, the flow value of being a challenger is just equal to the expected capital gain that results from making a *patentable* discovery, less the associated cash flow spent on R&D: $rV^C = \theta\lambda\sigma \cdot \left[V^I - V^C\right] - C(\sigma).$ The first order condition for the challenger's research intensity, holding constant the number of challengers, is simply $C'(\sigma) = \theta\lambda\left[V^I - V^C\right].$

The number of firms engaged in R&D competition is determined by the free entry condition, $V^C - k = 0$ (we assume there is no binding integer constraint). In the Appendix, we show that this implies $\theta\lambda\sigma\left[V^I - V^C\right] = C(\sigma) + rk.$ In other words, under free entry, the expected capital gain from making a *patentable* discovery is equal to cash flows required to produce it. A corollary result is that the average and marginal costs of obtaining a patentable invention are equal; that is $C'(\sigma)\sigma = C(\sigma) + rk.$

Of course V^I and V^C are also functions of σ and *n*. In the Appendix, we show that

[1] $$\left[V^I - V^C\right] = \frac{p\tilde{u} + C(\sigma)}{r + \theta\lambda(n+1)\sigma} = \frac{p\tilde{u} - rk}{r + \theta\lambda n\sigma}.$$

The difference between the values of being an incumbent and a challenger is simply the difference between the respective cash flows discounted by a measure of the *economic* life of patents. Note that the denominator in [1] is increasing in the arrival rate of patentable discoveries, $\theta\lambda n\sigma.$

From the preceding analysis, it is clear that we can characterize the R&D intensity of firms, $\sigma^*,$ and the equilibrium number of firms, *n**, using the following two expressions:

[2] $$C'(\sigma)\sigma = C(\sigma) + rk, \text{ and}$$

[3] $$C'(\sigma) = \frac{\theta\lambda\left[p\tilde{u} - rk\right]}{r + \theta\lambda n\sigma}.$$

3.2 Properties of the Equilibrium

In a stationary equilibrium, the firms that wish to compete will already have sunk their fixed R&D investments. If we consider marginal changes in certain parameters, the number of firms engaged in R&D would not decline because the expected value of actively competing in subsequent races is strictly positive (so long as $k > 0$).[12]

Instead, we compare two symmetric stationary equilibria involving industries that are identical except for the value of a single exogenous parameter. Firms take into account the exogenous parameters when deciding whether to incur the fixed cost of an R&D lab. In the Appendix we show the following:

> **Proposition 2** - (a) σ^* is independent of p and λ, and increasing in both r and k; (b) n^* is increasing in p and λ, and decreasing in both r and k; (c) $n^* \cdot \sigma^*$ is increasing in p and λ, and decreasing in both r and k.

Proposition 2 shows that if R&D is cheaper, or more productive, in one of our hypothetical industries, more firms will enter that industry and it will innovate more rapidly. Somewhat surprisingly, there is no difference across industries in the amount of R&D effort by individual firms. So many firms enter that any extra rents one would expect to earn (because of cheaper or more productive R&D) are simply dissipated.[13]

The results are more complicated when comparing industries that differ only in the discount rate or the fixed cost of establishing an R&D lab. All else equal, a more expensive R&D facility can be amortized only if the firm is able to earn more rents from patentable inventions. If

[12] More precisely, in this model, once firms decide to enter the industry, there are no shocks to firms' research productivity or costs that would imply any subsequent entry or exit.

those rents are indeed higher, the firms that enter will race harder—firm level R&D spending will be higher than in the other industry. But this equilibrium can be supported only when fewer firms enter in the first place.[14] In other words, higher barriers to entry discourage entry but also increase the rents earned by the firms that do enter. The net result, proven in the Appendix, is less innovation at the industry level.

This model can be compared to the one in Hunt (1999a) where the number of firms is exogenous. In a number of instances, the results for firm level R&D investments vary from those reported here. For example, in that model, more firms or more productive R&D may be associated with smaller R&D investments by individual firms. An increase in the relative price of R&D also reduces firm level R&D investments. Still, the comparative static results reported in that paper for rates of innovation at the industry level are the same as reported here.[15]

4. Patentability Standards and the Rate of Innovation

We typically think of the U.S. patent system as applying a common set of criteria to inventions in all technology fields and industries. In this section, however, we construct a hypothetical in which two otherwise identical industries are subject to different standards of patentability. Firms take this standard into account when deciding whether to sink the fixed cost of an R&D lab. In this way we allow for the possibility that patentability criteria affect the number of firms engaged in R&D.

---

[13] To see this, note that p and λ are absent in equation [2]. So long as the cost curve is weakly convex, for any given specification of r and k, the equation is satisfied for only one value of σ. This is accomplished by a difference in n in equation [3] that exactly offsets the effect of any difference in p or λ.

[14] Larger values of k increase the value of σ where equation [2] is satisfied. To maintain the equality in [3], n must take a smaller value. A similar intuition applies for larger values of r.

[15] In a model where there is free entry, but where firms must also sink fixed R&D costs in each patent race, the comparative static results at the firm level would lie somewhere between those reported here and in Hunt (1999a). The results for rates of innovation at the industry level are the same as reported here.

In equilibrium, firms equate the marginal cost of additional R&D effort to the expected gain associated with inventing first. This gain is affected by patentability criteria in two ways. First, there is the likelihood that any given invention by a firm qualifies for protection. Second, there is a relationship between this probability and the number of rivals a firm competes with. In the Appendix we show the following:

Proposition 3 - In the stationary symmetric equilibrium, (a) σ^* is independent of s; (b) there exists $s^* \in [0, \overline{u}]$ s.t. $\forall s \in [0, s^*)$, n^* is increasing in s and $\forall s \in (s^*, \overline{u}]$, n^* is decreasing in s; (c) s^* is increasing in p and λ, and decreasing in r and k.

Thus differences in the standard of patentability (the inventive step) do not affect the R&D intensity of individual firms, but they do affect the number of firms actively engaged in R&D and, therefore, the industry-wide rate of innovation. For a given set of parameters, there is a unique standard (s^*) that maximizes the number of firms engaged in R&D.

Part (c) of Proposition 3 is again based on a comparison of two industries identical in every respect except for one of the exogenous parameters that influence the rate of innovation. Without loss of generality, assume that industry *A* innovates more rapidly than industry *B*. From Proposition 2 we know that this could be because marginal R&D costs less or is more productive. Or it may be the case that in industry *A* either the discount rate or the fixed cost of establishing an R&D lab is smaller than in industry *B*. Regardless of the particular mechanism, Proposition 3 shows that the R&D maximizing standard of patentability is stricter for the industry that is predisposed to innovate rapidly.

If patentability standards were set for each industry, and the R&D maximizing standard was chosen in each case, a smaller proportion of innovations would qualify for protection in industry *A* than in industry *B*. Let s_A^* and s_B^* denote the critical patentability for industry *A* and *B*, respectively. Relative to this benchmark, if a common standard, $\tilde{s}$, were applied to both industries, neither industry would innovate as rapidly because fewer firms would enter these industries. Now suppose we consider a common standard that is less strict than $\tilde{s}$. If $\tilde{s}$ lies between s_A^* and s_B^*, there will be even less innovation in industry *A* and somewhat more innovation in industry *B*, because even fewer firms would enter industry *A* but some additional firms would enter industry *B*. The net effect, in terms of welfare, depends on whether such a change generates so much additional innovation in industry *B* that it offsets the lower innovation in industry *A*.[16]

Note that this result is characterized in terms of the rate of innovation of industries and not individual firms. If the only difference between our hypothetical industries is the fixed cost of setting up an R&D lab, individual firms in industry *B* may do more R&D than individual firms in industry *A*. Still, the R&D maximizing standard of patentability would be stricter in industry *A*. This suggests the need for caution in empirical work, as it is at least theoretically possible to erroneously associate weaker patentability standards with improved R&D incentives.

4.1 Deriving the Critical Patentability Standard

In the Appendix, we show that *s** is implicitly defined by the following equation:

$$\left(\frac{\theta\lambda(n+1)\sigma}{r+\theta\lambda(n+1)\sigma}\right)[p\tilde{u}+C(\sigma)]=[ps+C(\sigma)]. \tag{4}$$

As the standard of patentability is made more strict (requiring a larger inventive step),

[16] In examples where invention magnitudes are distributed normally and R&D costs are quadratic, the increase in innovation in industry *B* does not offset the larger decline in innovation that occurs in industry *A*. Of course, the welfare implications of such changes would depend on consumers' preferences between the two goods.

firms encounter the following tradeoff. On the one hand, a firm that makes a marginal discovery would not obtain a patent. The cost of this is the forgone value of the marginal patent plus the R&D expended in the subsequent patent race—the right-hand side of [4]. This is the static effect of an increase in the patentability standard.

But raising the standard also has a dynamic effect, because firms are able to earn flow profits for a longer period of time—the left-hand side of [4]. The expected gain is the average value of patentable inventions, plus the R&D that would otherwise be expended in the next patent race. But the gains enjoyed by incumbents occur at the end of their tenure and are discounted accordingly. Thus the relative weight placed on these static and dynamic effects depends on the industry-wide arrival rate of patentable discoveries.

It may seem counter-intuitive that the benefit to preserving an incumbent's rents is larger when patentable inventions are more frequent. Given that the expected duration of those rents is shorter, the present value of the rents might be relatively small. But changes in the standard of patentability induce marginal changes in the duration of those rents. In a rapidly innovating industry, the rents that are affected are earned relatively soon and therefore are not discounted very much. In an industry that innovates less rapidly, increasing the standard of patentability contributes additional rents, but they are earned far in the future and are discounted accordingly.

Now consider how the tradeoff between static and dynamic effects changes as we vary the patentability standard from a very low to a very high value. When the standard is very weak ($s = 0$), the static effect is irrelevant because rents earned on the marginal invention are too small to affect the participation decision.[17] In this range, adopting a stricter standard would increase the number of firms actively engaged in R&D. But eventually, as the standard is made increasingly

[17] This is true whenever $ps < rk$.

more strict, the dynamic effect becomes smaller while the static effect becomes larger.[18] When the patentability standard is very strict, the static effect dominates. For any given specification of exogenous parameters consistent with an interior equilibrium there is only one standard, or height of the inventive step, where the two effects are exactly equal.

4.2 The Socially Optimal Standard

Unlike firms, society enjoys the benefits of all innovations and enjoys those benefits forever. Given that the productivity and cost of R&D and the distribution of invention magnitudes do not change over time, the socially optimal R&D intensity and number of firms will be the same in every patent race. Social welfare is then

$$W(h,n) \equiv \frac{n}{r}\left\{\lambda h \frac{p\tilde{v}}{r} - C(h) - rk\right\}$$

where $\tilde{v} \equiv \int_{\underline{u}}^{\overline{u}} u dF(u)$ is the average value of inventions. In the Appendix we prove

> **Proposition 4**: Under the assumptions specified in Proposition 1, (a) the first best solution is $\sigma^{1B} > \sigma^*,\ \forall s \in [\underline{u},\ \overline{u}],$ and $n^{1B} = \infty;$ (b) the second best solution, where the social planner is limited to specifying the inventive step, is achieved by setting $s = s^*$; (c) $\sigma^{2B} = \sigma^*,\ n^{2B} = n^*(s^*)$.

An unconstrained social planner would set the R&D intensity of each firm so that the marginal cost of generating an innovation would just equal the expected social benefit: $\lambda p\tilde{v}/r$. That amount always exceeds the expected private return earned by firms in the model, so the first best R&D intensity is always larger than the R&D intensity observed in the private equilibrium.

---

[18] In other words, the industry-wide arrival rate of patentable inventions, $\theta\lambda n\sigma$, eventually declines as s increases.

Another property of the private equilibrium is that the expected benefit earned by the innovating firm just equals the expected cost of making the discovery. But if the expected social benefit exceeds the expected private benefit, it also exceeds the expected cost of making discoveries.[19] So the social planner would prefer that an infinite number of firms establish an R&D lab.

Now suppose that the social planner is limited to choosing the inventive step in order to maximize welfare. The second best maximization problem is then

$$W^{2\mathrm{B}} \equiv \underset{s}{Max}\left\{\frac{n(s)}{r}\left(\lambda\sigma(s)\frac{p\tilde{v}}{r}-C(\sigma(s))-rk\right)\right\},$$

where $\sigma(s)$ and $n(s)$ are the equilibrium research intensity and number of firms that arise in the equilibrium characterized in Proposition 1. From Proposition 3, we know that the social planner can maximize the number of firms that enter the industry by setting $s = s^*$. This, in turn, will achieve the most rapid rate of innovation that can be attained in the private equilibrium. But we also know from Proposition 3 that the social planner cannot influence the R&D intensity of the firms that enter.[20] Consequently, in the second best solution the number of active firms and the R&D intensity of those firms is strictly lower than their first best levels.

5. Discussion and Conclusions

This paper develops a model of cumulative innovation 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, patentability criteria, in particular the requirement of

---

[19] This follows from the free entry condition and the fact that the private and social costs of generating innovations are the same.

[20] In a model that assumes a fixed number of firms, the planner's choice of s would affect the R&D investments of individual firms and therefore the entire industry. The normative implications are the same. See Hunt (1999a).

nonobviousness (the inventive step) play an important role in determining the share of future discoveries that will affect the expected profits earned on patented inventions discovered today.

In such an environment, there exists a unique inventive step that maximizes the rate of innovation in an industry, by maximizing the number of firms that enter into R&D competition. The effect of changes in the inventive step on the industry-wide rate of innovation depends on whether the initial standard is more or less stringent than this critical value. This critical standard will be more stringent for industries predisposed to innovate rapidly than for industries predisposed to innovate slowly. In other words, under the R&D maximizing patentability standard, a smaller share of inventions should qualify for protection in rapidly innovating industries than in other industries.

When a common inventive step is applied to all industries in an economy, the number of firms engaged in R&D in each of those industries will depend on the stringency of the standard. Generally speaking, when the standard is more stringent, there will be more firms in the industries disposed to innovate more rapidly, and fewer firms in industries disposed to innovate less rapidly. The converse would be true under a weaker standard. In this context, there is an element of industrial policy involved in setting a common patentability standard.

5.1 International Implications

If an R&D maximizing patentability standard was set in different economies, with different mixes of industries, it's likely they would not be the same. The standard would likely be stricter in economies that enjoy a comparative advantage in R&D. Adopting the same standard for all countries may increase the rate of innovation in some countries but might reduce it in others. Of course, more general statements about welfare implications require a model that allows for trade, foreign direct investment, and licensing.[21]

Efforts toward patent harmonization have thus far concentrated on issues such as establishing uniform priority, a minimum patent length, fewer subject matter exceptions, adequate remedies for infringement (damages, injunctions), and adequate administrative and judicial infrastructures. One exception was the proposed Patent Harmonization Treaty, abandoned in the mid 1990s, which included a specification of patentability standards (Moy 1993). Recently, the U.S. Patent and Trademark Office proposed to include, among other things, an American-style nonobviousness test in its agenda for future international negotiations on patent harmonization (USPTO 2001).

5.2 A Common Law Standard of Patentability?

Given that the critical standard is a function of industry characteristics that influence the industry's rate of innovation, this standard will vary as those characteristics change. An economy-wide increase in the productivity of R&D, for example, might suggest the inventive step should be increased in order to obtain the maximum possible benefit of this new-found productivity. If the productivity increase occurred in a single industry, a social planner might adopt a more stringent standard, but doing so could reduce the rate of innovation in the other industries.

The critical inventive step derived from the model presented in this paper follows from an explicit balancing of the gains and losses generated by marginal changes in the patent standard. A social planner would reduce the inventive step until the value of granting exclusive rights to

[21] See the surveys by Maskus (2000) and Saggi (2002). Using the example of computer software, Weiss (2004) derives the changes in welfare that result from the sequential adoption of patent protection by different countries.

the marginal invention is just equal to the expected value of rents that are lost as the economic life of patents is reduced. This has the flavor of a common law balancing test rather than an explicit standard specified by law.

One can argue that for a very long time, that is how the requirement of nonobviousness functioned in the U.S. patent system. The requirement existed in court precedents about a century before it appeared in the 1952 Patent Act, which largely adopted the test used by the courts. The classic articulation of the test appeared in the 1966 decision *Graham v. John Deere*: At the time it was made, would the invention have been obvious to a practitioner of ordinary skill in the relevant field? If such a determination is influenced by factors such as research productivity or costs, the judicial test and the one described in this paper appear quite similar.

Recently, some legal scholars have argued that patent standards should be influenced by a balancing of costs and benefits (Barton 2001, Rai 2003). But many patent practitioners and scholars support relatively stable patentability criteria and an equal treatment of all patentable technologies. They argue the patent system is already costly and additional complexity would only increase these costs while also increasing uncertainty about future returns.

5.3 The American Policy Experiments of the 1980s and 1990s

But patent standards have been changed before. During the 1980s, the U.S. adopted a new form of intellectual property (mask rights) to protect the physical layout of semiconductor chips and a series of court decisions reduced the inventive step for patents (Hunt 1999a, 1999b). The patentability of computer programs was firmly established by the mid 1990s (Hunt 2001). Supporters of these changes argued they would stimulate innovation in America's high technology industries. The results of this paper suggest the opposite might well be true —

weaker patentability standards are more likely to increase R&D in industries that innovate slowly, and to reduce R&D in industries that would otherwise innovate rapidly.

The final assessment of the changes adopted in the 1980s remains an open empirical question.[22] This model suggests at least one testable implication: historical patterns of entry and exit from industries may have changed in some systematic way – with relatively more net entry into industries that innovate slowly and relatively less net entry into industries that innovate more rapidly. Testing that hypothesis is an important topic for future research.

[22] Bessen and Maskin (2002) and Bessen and Hunt (2004) argue that granting patents on computer software may have been detrimental. See also Kortum and Lerner (1999) and the reviews by Jaffe (2000) and Hunt (1999b).

REFERENCES

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

Barton, John H. 2001. "Nonobviousness," mimeo, Stanford University Law School.

Bernheim, Douglas B. 1984. "Strategic Deterrence of Sequential Entry into an Industry," *RAND Journal of Economics*. Vol. 15, pp. 1-11.

Bessen, James and Eric Maskin. 2002. "Sequential Innovation, Patents, and Imitation," mimeo, Research on Innovation.

Bessen, James and Robert M. Hunt. 2004. "An Empirical Look at Software Patents," Working Paper No. 03-17/R, Federal Reserve Bank of Philadelphia.

Cadot, Olivier and Lippman, Steven A. 1997. "Fighting Imitation with Fast Paced Innovation," INSEAD Working Paper No. 97-73.

Chou, Teyu and Hans Haller. 1995. "The Division of Profit in Sequential Innovation Reconsidered," mimeo, Virginia Polytechnic Institute and State University.

Cohen, Wesley M. and Richard C. Levin. 1989. "Empirical Studies of Innovation and Market Structure," in Schmalensee, Richard and Robert D. Willig, eds., *Handbook of Industrial Organization*. Amsterdam: North Holland, pp. 1059-1107.

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

Denicolò, Vincenzo. 2000. "Two-Stage Patent Races and Patent Policy," *RAND Journal of Economics*. Vol. 31, pp. 488-501.

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

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

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

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

Horowitz, Andrew W. and Edwin L.-C Lai. 1996. "Patent Length and the Rate of Innovation," *International Economic Review*. Vol. 37, pp. 785-801.

Hunt, Robert M. 2001. "You Can Patent That? Are Patents on Computer Programs and Business Methods Good for the New Economy?" Federal Reserve Bank of Philadelphia *Business Review* (1st Quarter), pp. 5-15.

REFERENCES

_____. 1999a. "Nonobviousness and the Incentive to Innovate: An Economic Analysis of Intellectual Property Reform," Working Paper 99-3, Federal Reserve Bank of Philadelphia.

_____. 1999b. "Patent Reform: A Mixed Blessing for the U.S. Economy?" Federal Reserve Bank of Philadelphia *Business Review* (November/December), pp. 15-29.

Jaffe, Adam B. 2000. "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process," *Research Policy*. Vol. 29, pp. 531-7.

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

Kortum, Samuel and Josh Lerner. 1999. "What Is Behind the Recent Surge in Patenting?" *Research Policy*. Vol. 28, pp. 1-22.

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

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

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

Maskus, Keith. E. 2000. *Intellectual Property Rights in the Global Economy*. Washington: Institute for International Economics.

Moy, R. Carl. 1993. "The History of the Patent Harmonization Treaty: Economic Self-Interest as an Influence," *The John Marshall Law Review*. Vol. 26, pp. 457-95.

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

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

O'Donoghue, Ted, Suzanne Scotchmer and Jacques-Francois Thisse. 1998. "Patent Breadth, Patent Life, and the Pace of Technological Progress," *Journal of Economics and Management Strategy*. Vol. 7, pp. 1-32.

Rai, Arti K. 2003. "Engaging Facts and Policy: A Multi-Institutional Approach to Patent System Reform," *Columbia Law Review*. Vol. 106, pp. 1035-1135.

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

REFERENCES

Saggi, Kamal. 2002. "Trade, Foreign Direct Investment, and International Technology Transfer: A Survey," *World Bank Research Observer*. Vol. 17, pp. 191-235.

Scherer, Frederic M. 1992. "Schumpeter and Plausible Capitalism," *Journal of Economic Literature*. Vol. XXX, pp. 1416-33.

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

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

U.S. Patent and Trademark Office. 2001. "Request for Comments on the International Effort to Harmonize the Substantive Requirements of Patent Laws," *Federal Register*, Vol. 66, pp. 15409-11.

Weiss, Pia. 2004. "Competitive Disadvantage Through Non-existing Software Patents," mimeo Chemnitz University of Technology.

APPENDIX

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

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

(ii) $C''(h) > 0\ \forall h > \hat{h} \in [0, \overline{h});$

(iii) $\text{Lim}_{h \to 0}\ C(h)/h = \text{Lim}_{h \to 0}\ C'(h) = 0;$

(iv) $\text{Lim}_{h \to \infty}\ C'(h) = \infty;$

(v) $C(\overline{h}) < \infty,\ \forall \overline{h} \in [0, \infty).$

Then, there exists a unique, stationary, symmetric equilibrium of the game, characterized by the pair (σ^*, n^*), where there are n^* challengers who choose a flow R&D intensity $\sigma^* \in (0, \overline{h}]$.

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

Lemma 1 - Suppose $V^W_{q+1} \in (0, \infty)$ and $V^W_{q+1} - V^L_{q+1} > 0.$ If rivalry and the fixed R&D costs are sufficiently small, at least one challenger will choose to enter a stage game.

Proof: Note that we are treating the continuation values as exogenous parameters. Later we show that, in equilibrium, the continuation values satisfy the requirements set out in the lemmas.

Consider the case where there is no rivalry and fixed R&D costs are zero. We need to show that $V^i_q(h^i, 0) \geq V^i_q(0, 0).$ The inequality is satisfied when there exists some positive level of R&D intensity where $\lambda h^i_q V^W_{q+1} \geq C(h^i_q),$ which is satisfied if the minimum average cost of R&D is not too high. The third assumption ensures there is at least one R&D intensity $\tilde{h} \in (0, \overline{h}]$ where the inequality is strict.

Now we consider a strictly positive fixed R&D cost k. In that case, a challenger chooses to enter so long as $\lambda h^i_q V^W_{q+1} \geq C(h^i_q) + [r + h^i_q]k.$ If $\lambda \hat{h}^i_q V^W_{q+1} > C(\hat{h}^i_q)$, there is also a level of fixed R&D cost where $\lambda \hat{h}^i_q V^W_{q+1} \geq C(\hat{h}^i_q) + [r + \hat{h}^i_q]k.$ Thus for k sufficiently small, there are always at least two firms, an incumbent and at least one challenger.

Now suppose there is some small positive level of rivalry. A challenger will enter the stage game if the following inequality holds:

$$\lambda h^i_q \left\{ r V^W_{q+1} + \lambda a^i_q \left[V^W_{q+1} - V^L_{q+1} \right] \right\} \geq \left[r + \lambda a^i_q \right] C(h^i_q) \left[r + \lambda h^i_q + \lambda a^i_q \right] rk. \quad \text{[A.1]}$$

Applying the preceding argument to this inequality, for k and a^i_q sufficiently small, there is an R&D intensity in the interval $(0, \overline{h}\,]$ where this inequality is strict. So long as $V^W_{q+1} > 0$ and $V^W_{k+1} - V^L_{k+1} > 0$, the magnitude of the continuation values will always define a set of pairs $(a^i_q, k) \in \mathbb{R}^+$ where the participation constraint is satisfied. We can also define a level of fixed R&D cost, $\hat{k}(a^i_q)$ where the participation constraint just binds.■

Lemma 2 - If $V^W_{q+1} \in (0, \infty)$, $V^W_{q+1} - V^L_{q+1} > 0,$ and $k < \hat{k}(0)$, there exists an interior equilibrium of the stage game.

Proof: The proof of existence is a modification of the existence proof in Reinganum (1985). We continue to treat the continuation values as exogenous parameters but take into account the effect of a firm's choice of R&D intensity on the likelihood of winning and the expected length of the patent race. Firms take their rivals' research intensity as given. Fixed R&D costs must be sufficiently low so that at least one firm is willing to engage in R&D.

The derivative of the firm's objective function, $\partial V^i_q / \partial h^i_q$, is

APPENDIX

[A.2] $$\frac{r\left[\lambda V^{W}_{q+1}-C'(h^{i}_{q})\right]+\lambda a^{i}_{q}\left[\lambda[V^{W}_{q+1}-V^{L}_{q+1}]-C'(h^{i}_{q})\right]+\lambda\left[C(h^{i}_{q})-C'(h^{i}_{q})h^{i}_{q}\right]}{\left[r+\lambda(h^{i}_{q}+a^{i}_{q})\right]^{2}}.$$

The sign of [A1] is the sign of the numerator, which we will call $\phi^{i}(h^{i}_{q},a^{i}_{q})$. Note that $\phi^{i}(h^{i}_{q},a^{i}_{q})$ is strictly decreasing in R&D intensity:

$$\frac{\partial\phi^{i}(h^{i}_{q},a^{i}_{q})}{\partial h^{i}_{q}}=-C''(h^{i}_{q})\left[r+\lambda(h^{i}_{q}+a^{i}_{q})\right]<0.$$

If the saturation point of R&D ($\overline{h}$) is sufficiently large, there will be a finite level of R&D effort where $\phi^{i}(h^{i}_{q},a^{i}_{q})=0.$ $V^{i}_{q}(h^{i}_{q},a^{i}_{q})$ is maximized by this level of R&D effort. Let $h^{i}_{q}(a^{i}_{q})$ denote the firm's best response to the level of rivalry it encounters. The strict monotonicity of $\phi^{i}(h^{i}_{q},a^{i}_{q})$ implies that this best response is unique. The strategy space is a convex, compact, nonempty subset of $\mathbb{R}^{n}$, denoted $X\equiv\prod_{i=1}^{n}[0,\overline{h}]$. The vector $[h^{1}_{q}(a^{1}_{q}),h^{2}_{q}(a^{2}_{q}),\ldots h^{n}_{q}(a^{n}_{q})]$ maps X into itself continuously. Existence of an equilibrium then follows from Brouwer's fixed point theorem. ■

Lemma 3 - If $\lambda[V^{W}_{q+1}-V^{L}_{q+1}]-[C'(h_{q})+h_{q}C''(h_{q})]<0$, there exists a unique, symmetric equilibrium of the stage game.

Proof: Existence of a symmetric equilibrium follows from the firm's objective function and first order condition, which varies only by the level of rivalry encountered. In the symmetric equilibrium, $\phi^{i}(h^{i}_{q},a^{i}_{q})$ becomes $\phi^{i}\left(h_{q},(n-1)h_{q}\right)$. The corresponding first order condition is

$$r\left[\lambda V^{w}_{q+1}-C'(h_{q})\right]+\lambda(n-1)h_{q}\left[\lambda[V^{w}_{q+1}-V^{l}_{q+1}]-C'(h_{q})\right]+\lambda\left[C(h_{q})-C'(h_{q})h_{q}\right]=0.$$

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

$$\lambda\left[V^{w}_{q+1}-V^{l}_{q+1}\right]-\left[C'(h_{q})+h_{q}C''(h_{q})\right]<0.$$ ■

The symmetric equilibrium R&D intensity of the stage game with continuation values V^{W}_{q+1} and V^{L}_{q+1} is denoted $h_{q}(V^{W}_{q+1},V^{L}_{q+1})$.

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 (1991), 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 $p\cdot\overline{u}/r$. Per period returns for challengers are contained in the interval $[-C(\overline{h})/(r+\overline{h}),0]$. ■

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:

[A.3] $$C'(h_{q})\left[r+\lambda nh_{q}\right]=\lambda\left[rV^{W}+\lambda(n-1)h_{q}[V^{W}-V^{L}]+C(h_{q})\right].$$

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 in each race. Lemma 3 establishes the existence of such a best response for a given specification of the continuation values.

Note that $V^{W} = \theta V^{I} + (1-\theta)V^{C}$ and $V^{L} = V^{C}$. These continuation values take a simple recursive form:

$$V^{I}(h) = \frac{p\tilde{u} + \lambda nh\left[\theta V^{C}(h) + (1-\theta)V^{I}(h)\right]}{r+\lambda nh} = \frac{p\tilde{u} + \theta\lambda nhV^{C}(h)}{r+\theta\lambda nh}, \text{ and}$$

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

Solving for $V^{I}(h)$ and $V^{C}(h)$, we have,

$$V^{I}(h) = \frac{[r+\theta\lambda h]p\tilde{u} - \theta\lambda nhC(h)}{r[r+\theta\lambda(n+1)h]} \qquad \text{and} \qquad V^{C}(h) = \frac{\theta\lambda hp\tilde{u} - \left[r+\theta\lambda nh\right]C(h)}{r[r+\theta\lambda(n+1)h]}.$$

If we substitute for $V^{I}(h)$ and $V^{C}(h)$ in equation [A.3], the first order condition reduces to

$$\text{[A.4]} \qquad C'(h) = \theta\lambda\left(\frac{p\tilde{u} + C(h)}{r+\theta\lambda(n+1)h}\right).$$

We use σ to denote the equilibrium R&D intensity that satisfies equation [A.4]. It can be verified that, using equation [A.4], the condition required in lemma 3 for the uniqueness of the symmetric equilibrium of the stage game is satisfied.

If we substitute for $V^{I}(h)$ and $V^{C}(h)$ in equation [A.1], the participation constraint is simply $V^{C}(h) \geq k$. This in turn implies $C(h) + rk \leq \theta\lambda h\left[V^{I}(h) - V^{C}(h)\right] = C'(h)h.$ When the participation constraint binds, we can also express the first order condition as

$$\text{[A.5]} \qquad C'(h) = \theta\lambda\left(\frac{p\tilde{u} - rk}{r+\theta\lambda nh}\right).$$

During each race, for every challenger, the R&D intensity σ is the unique best response to the continuation values $V^{I}(\sigma)$ and $V^{C}(\sigma)$. 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. 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 (1991), 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 [A.4]. At $h = 0$, $C'(h) = 0$ while $\theta\lambda[V^{I}(h) - V^{C}(h)] = \theta\lambda p\tilde{u}/r$. Thus at the first intersection, the marginal cost curve must be rising faster than $\theta\lambda[V^{I}(h) - V^{C}(h)]$. If we can rule out an intersection where $\theta\lambda[V^{I}(h) - V^{C}(h)]$ is rising faster than marginal cost, we are done. Define $\mathrm{M}^{1} = C'(h)\left[r+\theta\lambda(n+1)h\right] - \theta\lambda[p\tilde{u} + C(h)] = 0$ and note that:

$$\text{[A.6]} \qquad \frac{\partial \mathrm{M}^{1}}{\partial h} = C''(h)\left[r+\theta\lambda(n+1)h\right] + C'(h)\theta\lambda n > 0.$$

This rules out an intersection where the marginal cost curve crosses $\theta\lambda[V^{I}(h) - V^{C}(h)]$ from above. ■

Proposition 2 - (a) σ* is independent of *p* and λ, and increasing in both *r* and *k*; (b) *n** is increasing in *p* and λ, and decreasing in both *r* and *k*; (c) *n**·σ* is increasing in *p* and λ, and decreasing in both *r* and *k*.

Proof: We reintroduce the relative price of outputs in terms of inputs (*p*) and rewrite [A.4] and [A.5] in the following form:

APPENDIX

$$M \equiv \begin{array}{l} \mathrm{M}^1 = C'(\sigma)\sigma - C(\sigma) - rk = 0 \\ \mathrm{M}^2 = C'(\sigma)[r + \theta\lambda n\sigma] - \theta\lambda[p\tilde{u} - rk] = 0 \end{array}$$

We'll need the following derivatives:

$$\begin{array}{ll} \mathrm{M}^1_\sigma = C''(\sigma)\sigma & \mathrm{M}^2_\sigma = C''(\sigma)\left[r + \theta\lambda n\sigma\right] + \theta\lambda n C'(\sigma) \\ \mathrm{M}^1_n = 0 & \mathrm{M}^2_n = C'(\sigma)\theta\lambda\sigma \\ \mathrm{M}^1_p = 0 & \mathrm{M}^2_p = -\theta\lambda\tilde{u} \\ \mathrm{M}^1_\lambda = 0 & \mathrm{M}^2_\lambda = -\dfrac{r}{\lambda}C'(\sigma) \\ \mathrm{M}^1_r = -k & \mathrm{M}^2_r = C'(\sigma) + \theta\lambda k \\ \mathrm{M}^1_k = -r & \mathrm{M}^2_k = \theta\lambda r \end{array}$$

The Jacobian $\left|\mathrm{M}\right| = \mathrm{M}^1_\sigma \mathrm{M}^2_n - \mathrm{M}^2_\sigma \mathrm{M}^1_n = C''(\sigma)C'(\sigma)\theta\lambda\sigma^2 > 0.$

i. Increasing the output price:

$$\frac{\partial\sigma}{\partial p} = \frac{\mathrm{M}^2_p \mathrm{M}^1_n - \mathrm{M}^1_p \mathrm{M}^2_n}{\left|\mathrm{M}\right|} = 0;$$

$$\frac{\partial n}{\partial p} = \frac{\mathrm{M}^1_p \mathrm{M}^2_\sigma - \mathrm{M}^2_p \mathrm{M}^1_\sigma}{\left|\mathrm{M}\right|} = \frac{\tilde{u}}{C'(\sigma)\sigma} > 0.$$

ii. Increasing the productivity of R&D:

$$\frac{\partial\sigma}{\partial \lambda} = \frac{\mathrm{M}^2_\lambda \mathrm{M}^1_n - \mathrm{M}^1_\lambda \mathrm{M}^2_n}{\left|\mathrm{M}\right|} = 0;$$

$$\frac{\partial n}{\partial \lambda} = \frac{\mathrm{M}^1_\lambda \mathrm{M}^2_\sigma - \mathrm{M}^2_\lambda \mathrm{M}^1_\sigma}{\left|\mathrm{M}\right|} = \frac{r}{\theta\lambda^2\sigma} > 0.$$

iii. Increasing the fixed cost of setting up an R&D lab:

$$\frac{\partial\sigma}{\partial k} = \frac{\mathrm{M}^2_k \mathrm{M}^1_n - \mathrm{M}^1_k \mathrm{M}^2_n}{\left|\mathrm{M}\right|} = \frac{r}{C''(\sigma)\sigma} > 0;$$

$$\frac{\partial n}{\partial k} = \frac{\mathrm{M}^1_k \mathrm{M}^2_\sigma - \mathrm{M}^2_k \mathrm{M}^1_\sigma}{\left|\mathrm{M}\right|} = \frac{-r\left[C''(\sigma)\left[r + \theta\lambda(n+1)\sigma\right] + C'(\sigma)\theta\lambda n\right]}{C''(\sigma)C'(\sigma)\theta\lambda\sigma^2} < 0.$$

The change in industry-wide R&D is therefore:

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

iv. Increasing the discount rate:

$$\frac{\partial\sigma}{\partial r} = \frac{\mathrm{M}^2_r \mathrm{M}^1_n - \mathrm{M}^1_r \mathrm{M}^2_n}{\left|\mathrm{M}\right|} = \frac{k}{C''(\sigma)\sigma} > 0;$$

$$\frac{\partial n}{\partial r} = \frac{\mathrm{M}^1_r \mathrm{M}^2_\sigma - \mathrm{M}^2_r \mathrm{M}^1_\sigma}{\left|\mathrm{M}\right|} = -\left(\frac{C''(\sigma)C'(\sigma)\sigma + C''(\sigma)\left[r + \theta\lambda(n+1)\sigma\right]k + C'(\sigma)\theta\lambda nk}{C''(\sigma)C'(\sigma)\theta\lambda\sigma^2}\right) < 0.$$

The change in industry-wide R&D is therefore:

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

Proposition 3 - In the stationary symmetric equilibrium, (a) σ* is independent of *s*; (b) there exists $s^*\in[0,\bar{u}]$ s.t. $\forall s\in[0,s^*)$, *n** is increasing in *s* and $\forall s\in(s^*,\bar{u}]$**,** *n** is decreasing in *s*; (c) *s** is increasing in *p* and λ, and decreasing in *r* and *k*.

Proof: We begin by calculating the derivatives**:**

$$\mathrm{M}_s^1=0;$$
$$\mathrm{M}_s^2=\lambda\left\{\frac{\partial\theta}{\partial s}\left[C'(\sigma)n\sigma+rk\right]-\frac{\partial\left(\theta\tilde{u}\right)}{\partial s}p\right\}.$$

Recall that $\theta=1-F(s)$ and $\theta\tilde{u}=\int_s^{\bar{u}}udF(u),$ which implies that $\mathrm{M}_s^2=-\lambda f(s)\Psi(s),$ where

[A.7] $$\Psi(s)=\left\{\left[C'(\sigma)n\sigma+rk\right]-ps\right\}.$$

The expression used in the text is obtained by substituting the first order and free entry conditions for $C'(\sigma)$ and *rk* respectively. The comparative static calculations are thus:

$$\frac{\partial\sigma}{\partial s}=\frac{\mathrm{M}_s^2\mathrm{M}_n^1-\mathrm{M}_s^1\mathrm{M}_n^2}{\left|\mathrm{M}\right|}=0;$$
$$\frac{\partial n}{\partial s}=\frac{\mathrm{M}_s^1\mathrm{M}_\sigma^2-\mathrm{M}_s^2\mathrm{M}_\sigma^1}{\left|\mathrm{M}\right|}=\frac{f(s)\Psi(s)}{C'(\sigma)\theta\sigma}.$$

This establishes part (a) of the proof. For part (b), we must calculate the slope of $\Psi(s)$:

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

If there is a value $s^*\in[0,\bar{u}],$ where Ψ(*s**) = 0, we know that $\partial\Psi(s)/\partial s\big|_{s^*}=-p.$ Thus there can be at most one extremum of Ψ(*s*). Next we check the values of Ψ(*s*) as $s\to\bar{u}$ and $s\to 0$. These are evaluated most easily if we use [A.5] to substitute for $C'(\sigma)$ in [A.7] and evaluate

$$\Psi(s)=\left(\frac{\theta\lambda n\sigma}{r+\theta\lambda n\sigma}\right)\left[p\tilde{u}(s)-rk\right]-\left[ps-rk\right].$$

$$\mathrm{Lim}_{s\to\bar{u}}\Psi(s)=\left(\frac{0\cdot\lambda n(\bar{u})\sigma(\bar{u})}{r+0\cdot\lambda n(\bar{u})\sigma(\bar{u})}\right)\left[p\bar{u}-rk\right]-\left[p\bar{u}-rk\right]<0.$$

If the participation constraint is satisfied when *s* = 0 (i.e. $p\tilde{u}(0)-rk\geq 0$), the second limit is:

$$\mathrm{Lim}_{s\to 0}\Psi(s)=\left(\frac{\lambda n(0)\sigma(0)}{r+\lambda n(0)\sigma(0)}\right)\left[p\tilde{u}(0)-rk\right]+rk>0.$$

But it is possible that, depending on the distribution of invention magnitudes and the output price, under a very weak patentability standard, the participation constraint [A.5] might be violated (i.e. $p\tilde{u}(0)-rk<0$). In that case $\Psi(s)$ does not exist at $s = 0$. Instead, define $\hat{s}$ s.t. $p\tilde{u}(\hat{s})-rk=0$. Then $\Psi(s)$ exists for $\forall\ s\in(\hat{s},\overline{u}]$. We also know that $\Psi(s)$ is initially positive for values of s just greater than $\hat{s}$, because $-[p\hat{s}-rk]>0$ implies that

$$\left(\frac{[1-F(\hat{s})]\lambda n(\hat{s})\sigma(\hat{s})}{r+[1-F(\hat{s})]\lambda n(\hat{s})\sigma(\hat{s})}\right)\left[p\tilde{u}(\hat{s})-rk\right]-\left[p\hat{s}-rk\right]>0.$$

Existence of the extremum then follows from continuity of $\Psi(s)$ over $(\hat{s},\overline{u}]$. This establishes part (b) of the proof.

For part (c), we compute derivatives of the implicit function $\Psi(s^*) = 0$ with respect to the exogenous parameters explored in Proposition 2:

[A.8] $$\frac{\partial s^*}{\partial z}=\frac{-\partial\Psi(s)}{\partial z}\Big/\frac{\partial\Psi(s)}{\partial s}=\frac{\partial\Psi(s)}{\partial z}\frac{1}{p},$$

where z is either p, λ, r, or k. Note also that

$$\frac{\partial\Psi(s)}{\partial z}=\frac{\partial\sigma}{\partial z}\left[C''(\sigma)\sigma+C'(\sigma)\right]n+\frac{\partial n}{\partial z}C'(\sigma)\sigma-\frac{\partial\left[ps-rk\right]}{\partial z}.$$

i. Higher output prices:

$$\frac{\partial s^*}{\partial p}=\frac{\tilde{u}(s^*)-s^*}{p}>0.$$

ii. More productive R&D:

$$\frac{\partial s^*}{\partial\lambda}=\frac{rC'(\sigma)}{p\theta\lambda^2}>0.$$

iii. Higher fixed R&D costs:

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

iv. A higher discount rate:

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

Proposition 4: Under the assumptions specified in Proposition 1, (a) the first best solution is $\sigma^{1B}>\sigma^*,\ \forall s\in[\underline{u},\ \overline{u}],$ and $n^{1B}=\infty$; (b) the second best solution, where the social planner is limited to specifying the inventive step, is achieved by setting $s = s^*$; (c) $\sigma^{2B}=\sigma^*,\ n^{2B}=n^*(s^*)$.

Proof: The social welfare function is simply

[A.9] $$W(h,n)\equiv\sum_{t=1}^{\infty}\left(\frac{\lambda nh}{r+\lambda nh}\right)^t\frac{p\tilde{v}}{r}-\sum_{t=0}^{\infty}\left(\frac{\lambda nh}{r+\lambda nh}\right)^t\frac{nC(h)}{r+\lambda nh}-nk=\frac{n}{r}\left\{\lambda h\frac{p\tilde{v}}{r}-C(h)-rk\right\},$$

where $\tilde{v}\equiv\int_{\underline{u}}^{\overline{u}}udF(u)$. The first derivatives with respect to h and n are

$$\frac{n}{r}\left\{\lambda\frac{p\tilde{v}}{r}-C'(h)\right\} \text{ and } \frac{1}{r}\left\{\lambda h\frac{p\tilde{v}}{r}-C(h)-rk\right\},$$

respectively. The first best R&D intensity σ^{1B} is a finite value that solves the equation $\lambda p\tilde{v}/r - C'(h) = 0.$ Note that

$$\lambda\frac{p\tilde{v}}{r} > \theta\lambda\left(\frac{p\tilde{u}-rk}{r+\theta\lambda nh}\right), \forall s,k \qquad \text{[A.10]}$$

which implies that $\sigma^{1B} > \sigma^*$. To see that the first best number of firms is infinite, recall that the binding free entry constraint in the private equilibrium implies $C'(\sigma^*)\sigma^* - C(\sigma^*) - rk = 0.$ But under the assumptions specified for the cost function in Proposition 1, this implies that the marginal cost of doing R&D exceeds the average cost at the first best R&D intensity. Thus $\lambda\sigma^{1B} p\tilde{v}/r - C(\sigma^{1B}) - rk > 0.$ Since the expected social benefit of innovations strictly exceed the expected cost, the social planner would prefer that an infinite number of firms establish an R&D lab and engage in R&D at the rate σ^{1B}. This establishes part (a) of the proposition.

The second best welfare maximization problem is simply

$$W^{2B} \equiv \underset{s}{Max}\left\{\frac{n(s)}{r}\left(\lambda\sigma(s)\frac{p\tilde{v}}{r}-C(\sigma(s))-rk\right)\right\}.$$

The first derivative is

$$\frac{\partial W}{\partial s} = \frac{\partial n(s)}{\partial s}\frac{1}{r}\left(\lambda\sigma(s)\frac{p\tilde{v}}{r}-C(\sigma(s))-rk\right)+\frac{\partial\sigma(s)}{\partial s}\frac{n(s)}{r}\left(\lambda\frac{p\tilde{v}}{r}-C'(\sigma(s))\right).$$

We know from Proposition 3 that $\partial\sigma/\partial s = 0.$ To show that W^{2B} is maximized at $s = s^*$, we need only show that $\lambda\sigma^* p\tilde{v}/r - C(\sigma^*) - rk > 0,$ which follows from [A.10]. This establishes parts (b) and (c) of the proposition.■