The Decision to Patent, Cumulative Innovation, and Optimal Policy*

Nisvan Erkal†
University of Melbourne

First version: October 2001
This version: October 2004

Abstract

This paper analyzes optimal policy in the context of cumulative innovation in a model that endogenizes patenting decisions of early innovators. Secrecy can significantly decrease investment in the second innovation. We show that as the effectiveness of secrecy as a protection mechanism increases, which may be the case if the government has a strong trade secret policy or innovators can monitor the flow of their technological information, it becomes optimal to have broad patent protection over a larger parameter space. In cases when patent policy is ineffective in achieving disclosure, it is socially desirable to have a lenient antitrust policy and allow collusive agreements.

JEL Classification: O31, O38, L40

Keywords: cumulative innovation, patenting decision, patent policy, antitrust policy

*This paper is based on the second chapter of my dissertation (University of Maryland, College Park, 2002). I am grateful to Rachel Kranton, Dan Vincent and Larry Ausubel for their valuable comments. I also thank David Brennan, John Creedy, Stephen King, Stephen Martin, two anonymous referees, and seminar participants at Australian Economic Theory Workshop (2004), European Economic Association Annual Congress (2003), the North American Summer Meeting of the Econometric Society (2003), the Intellectual Property Research Institute of Australia (IPRIA), and La Trobe University for their input. All errors are my own.

†Department of Economics, University of Melbourne, Victoria 3010, Australia. E-mail: n.erkal@unimelb.edu.au.

1 Introduction

Recent developments in the economic theory of innovation have emphasized its cumulative nature. One of the main issues addressed in the cumulative innovation literature is how patent and antitrust policy can be designed to divide the profits between sequential innovators in a way that provides them with optimal incentives to invest. Since the cumulative nature of innovation implies that the social value of innovations should include the value of subsequent innovations they inspire, several papers have argued for broad patent protection in order to provide early innovators with sufficient incentives to invest.[1]

One of the common assumptions made in this literature is that all innovations are patented. This paper challenges this assumption by pointing out that innovators frequently rely on secrecy to protect their innovations (Cohen et al., 2000; Levin et al., 1987).[2] When early innovators prefer secrecy over patenting, they can severely affect the investment incentives of rival firms in subsequent R&D races. The non-disclosure of early innovations also delays the benefit society gets from the use of those innovations. Hence, we argue that as the effectiveness of secrecy as a protection mechanism increases, the design of optimal policy should pay attention to encouraging disclosure and stimulating investment in subsequent innovations. This implies that it becomes optimal to have broad patent protection over a larger parameter space. If patent policy is ineffective in achieving disclosure, it is optimal to have a lenient antitrust policy.

[1] See Kitch (1977), Merges and Nelson (1990), Scotchmer (1991 and 1996), Green and Scotchmer (1995), Chang (1995), and Matutes et al. (1996). One exception is Denicolo (2000). He shows that having broad protection may not always be socially optimal if investment in each innovation is modeled in terms of a patent race.

[2] Cohen et al. (2000) find that patents tend to be the least preferred protection mechanism by firms while secrecy and lead time tend to be the most heavily used ones. They report that, especially in case of product innovations, firms use secrecy to protect just over 50% of their innovations. Moreover, by comparing their results with those of Levin et al. (1987), they conclude that there is an apparent growth in the importance of secrecy as an appropriability mechanism and a decline in the importance of patents. The importance of secrecy is further supported by Lerner (1994), who finds that 43% of all intellectual property litigation cases involve trade secrets.

We consider two consecutive R&D races as in Denicolo (2000). The firms race to achieve two symmetric and competing innovations, and the winner of the first race can also participate in the second race. At the end of the first race, the winner decides whether to patent the innovation or keep it secret. Under the patent law, an applicant is required to disclose sufficient information about the innovation to enable someone skilled in the art to make and use all the embodiments of the innovation claimed in the patent. Thus, the winner of the first race may be reluctant to patent, in order to get a head start in the second R&D race. After observing the patenting decision of the innovator, firms invest to develop an improved version in the second R&D race. If the first innovation is not disclosed, rival firms must spend resources to gain information about it or re-invent it, which reduces their competitiveness in the second R&D race.[3]

The two goals of the patent system are to encourage research and development, and to promote the disclosure of innovations. The analysis of optimal patent breadth has focused mainly on the first goal while our model aims to draw attention to the second. If innovators choose patenting, the effective life of a patent is determined by how the courts interpret the patent and antitrust laws.[4] If innovators choose secrecy, the government's trade secret policy determines the amount of protection innovators get against theft and unauthorized disclosure of their trade secrets. In this paper, taking trade secret policy as given, we analyze optimal patent and antitrust policy.

[3]If the innovator chooses to keep the innovation a secret, one issue that arises is what happens if the innovation is independently re-invented. In the United States, ownership is determined by the first-to-invent rule, which is what we assume in this paper. In most of the other countries, the first-to-file rule applies. We discuss briefly in Section 6 how a switch to the first-to-file rule may affect optimal policy.

[4]Patent policy and antitrust policy are interrelated since the ways in which patents are exploited by firms may give rise to antitrust concerns. For example, competing firms may use horizontal agreements regarding patents in order to restrict the prices they are going to charge in the product market. Or, they can give monopoly power to a single firm by selling all their competing patents to that firm. Historically, finding the right balance between the two areas of law has been an important policy question. See, for example, Tom (1998) and Pitofsky (2000) for comments on the interaction between intellectual property protection and antitrust law. See also the former Deputy Assistant Attorney General Joseph Farrell's speech, "Thoughts on Antitrust and Innovation," January 25, 2001, available at http://www.usdoj.gov/atr/public/speeches/7402.htm.

Specifically, following Chang (1995), we compare the behavioral implications and welfare properties of three policy regimes. The first regime is broad patent protection, where the courts find that the second innovation infringes the patent of the first innovation. Such a policy regime allows the first innovator to collect some licensing fees from the second innovator if it loses the second race. Under narrow patent protection, two possibilities exist; the antitrust authorities can have a policy of collusion or no collusion. Thus, we ask whether it may be desirable to permit collusive licensing deals between holders of competing but non-infringing patents. Such collusive agreements increases the investment incentives of innovators by eliminating the competition between them.

Taking into account the patenting incentives of early innovators suggests new ways in which broad patent protection or collusive licensing agreements may be socially desirable. The results illustrate that broad patent protection gives early innovators the highest incentives to patent. Thus, the desirability of broad patent protection depends on how attractive secrecy is for innovators. We find that if innovators cannot rely on secrecy to protect their innovations, which may be the case if the government has a weak trade secret policy or if innovators cannot monitor the flow of their technological information, it may not be optimal to have strong patent protection. This is because as the rival firm's probability of success in case of no patenting increases, innovating firms have increased incentives to patent their innovations. In such cases, early innovations are likely to be disclosed through patenting whether or not there is broad patent protection. Therefore, the government may prefer to have narrow patent protection in order to encourage investment in the second R&D race. If, on the other hand, early innovations are likely to be kept secret in the absence of broad patent protection, the government may find it optimal to have broad patent protection in order to encourage patenting. In such cases, the rivals' probability of success in case of no patenting is low and innovators can rely on secrecy as a protection mechanism. Having broad patent protection increases social welfare by encouraging earlier disclosure of innovations.

If early innovators cannot be encouraged to disclose their innovations under any of the policy regimes, it is never optimal to have broad patent protection. In such cases, it becomes optimal to allow collusive licensing agreements between competing innovators in order to stimulate investment in subsequent R&D races. This is because non-disclosure of early innovations decreases the investment incentives of second-generation investors and second-generation investors have the strongest incentives to invest when the courts find no infringement and allow collusion. That is, while broad patent protection encourages investment in the first race, collusion encourages investment in the second race. This is because under broad patent protection, the innovators end up sharing the value of the second innovation while under no infringement and collusion, they end up sharing the value of the first innovation.

Thus, the paper makes a case for having a lenient antitrust policy in industries where innovators rely on secrecy. This result differs from Chang (1995), Priest (1977) and Kaplow (1984), who find little or no support for collusion, and has important policy implications. The general policy of the US Department of Justice and the US Federal Trade Commission is to allow cross-licensing agreements and patent pools that include complementary patents, but prohibit those that include substitute patents (Shapiro, 2001). Analyzing the optimality of collusive licensing agreements in a dynamic setting where secrecy is a possibility, this paper shows that having a lenient antitrust policy may also be desirable in case of substitute patents.

The paper proceeds as follows. Section 2 presents the model and discusses the assumptions made regarding the legal background. In Sections 3 and 4, we analyze the private incentives of the firms and the optimal intellectual property policy respectively. Section 5 contains a detailed discussion of how our results differ from those in the literature. Section 6 concludes and indicates directions for future research. All proofs are in the Appendix.

2 The model

2.1 Research environment and consumers

The research environment is adapted from Denicolo (2000). There are two sequential R&D races and free entry into each race. The size of each innovation is exogenously given and commonly known, but its timing is stochastic. For simplicity, we assume the two innovations are symmetric in terms of both their private and social values. The winner of the first race can participate in the second race, so the model allows for repeated innovation by the same firm. If the winner of the second race is different from the winner of the first race, the second innovator can obtain a patent for the technology and start using it provided either that the new technology does not infringe the patent of the old technology or the firms are able to make a licensing deal.[5]

The R&D game is modeled using a Poisson discovery process. Following Loury (1979), it is assumed that at the beginning of each race, participant i pays a lump sum amount equal to cx_i, where c is the per-unit R&D cost and x_i is the R&D effort level chosen by the firm. We assume the hazard function is linear. Both the Poisson and the linear hazard function assumptions are made in order to make the model tractable. The common social and private discount rate is represented by r.

Once an R&D race ends and the innovation is successfully developed, the good can be produced at zero marginal cost. There is a mass of consumers with homogeneous preferences. They buy at most one unit of the good per time period. We assume a model of successive market domination similar to Scotchmer and Green (1990). Each innovation has a marginal market value of v per time period. Thus, the consumers are willing to pay v for the basic

[5] Whether the second innovation should be patentable or not is a policy question that we do not consider in this paper. We assume the second innovation is always patentable. In cases when it is not, the bargaining powers of second-generation innovators may be reduced. See Scotchmer and Green (1990), Scotchmer (1996), Denicolo (2000), and Denicolo and Zanchettin (2002) on optimal policy as far as the novelty requirement is concerned.

version and $2v$ for the improved version of the product per period. Firms compete by setting prices. If two different firms win the two races and the firms have to compete, the profits are given by 0 and v respectively. This is because Bertrand competition drives the price of the basic version to zero and that of the improved version to v minus some infinitesimal amount. Only the improved version will be sold in equilibrium. If the same firm is the owner of both of the innovations or if the firms sign a collusive licensing agreement, the profit is $2v$ per period.

2.2 Timing

The sequence of events is as follows. At the beginning of the first R&D race, after the government announces its intellectual property protection policy, firms simultaneously decide how much to invest in order to develop the first innovation. The race ends when one of the participants successfully develops the innovation. The winner of the race decides whether to patent and whether to commercialize the innovation.[6] In accordance with the goals of the patent system, we assume patenting means immediate full disclosure.[7] In other words, once an innovation is patented, the rival firms in the industry can easily build on it.[8] Therefore,

[6] In this paper, we assume patents have infinite duration and analyze the effect of patent breadth on innovators' disclosure incentives. Innovators' disclosure incentives could also be analyzed in a model where patent length is the critical policy instrument.

[7] For simplicity, we assume that if an innovator decides to apply for a patent, it is automatically given and the patent application is immediately disclosed to the public. In a more detailed analysis of the patenting process, it is necessary to differentiate between the time the innovator files a patent application, the time it becomes publicly available, and the time the patent is granted, if at all. Patent applications may sometimes be rejected if the authorities decide that they do not meet the requirements of novelty, usefulness, or nonobviousness. The uncertainty that characterizes the patenting process can be introduced into the model by assuming that when the innovator applies for a patent, it gets it with probability p. An interesting policy question would then be to investigate what the optimal p^* is. The magnitude of p would have a significant impact on the patenting decision of the innovator if there is pre-grant disclosure of patent applications. See Aoki and Spiegel (2003) on the impact of p on the patenting and R&D decisions of innovators in a model that takes into account pre-grant disclosure of patent applications.

[8] We assume that even if an innovation is not patented, rival firms in the industry learn about its achievement with no time lag. According to Mansfield (1985), information about a new product or a process leaks out within about a year.

if there is patenting, the hazard rates of the firms in the second R&D race are given by λx_w and λx_i, where x_w represents the investment of the winner of the first race in the second race. If the innovation is not patented but is commercialized, public disclosure of innovations significantly restricts the patenting rights of the original innovators. In the United States, innovators have a one-year grace period during which they can apply for a patent after they publicly disclose their innovations in any way. We assume that the length of the grace period is negligible compared with the expected duration of the R&D races. We also assume that reverse engineering is easy and that commercial use results in the same amount of information leakage as patenting. This assumption implies that if an innovator does not patent an innovation but offers it for sale, the rival firms have the same hazard rate as in the case of patenting.

If the first innovator does not disclose the innovation (through patenting or commercialization) and continues to work on it, the first-to-invent rule in the United States allows the innovator to claim ownership in case of independent re-invention.[9] In this case, the innovator's hazard rate is λx_w while the rival firms' hazard rate is μx_i, where $\lambda \geq \mu$. If the innovation is kept secret, the magnitude of μ, which is the marginal benefit of the rival firms' R&D efforts in case of non-disclosure, depends on the level of R&D spillovers. Such spillovers may take place through means such as employee mobility, informal communications between the employees of the firms, and espionage. Two factors determine whether innovators can protect themselves against such unauthorized disclosure: the strength of the existing trade secret policy and the effectiveness of their own monitoring efforts.

After the innovator makes the patenting and commercialization decisions, firms simulta-

[9] Under 35 U.S.C. § 102(g), a person is entitled to a patent unless "before such person's invention thereof, the invention was made in this country by another inventor who had not abandoned, suppressed, or concealed it." When an innovation should be regarded as abandoned, suppressed, or concealed is an issue that the courts have not been able to define with any level of precision. The prevailing view seems to be that if an innovator can prove that it is the first inventor and continues to work on the innovation even though it does not patent it, the innovation should not be considered abandoned, suppressed, or concealed.

neously choose how much to invest in the second R&D race. At the end of the second R&D race, the winner always patents since we assume the innovation is patentable and the winner does not lose anything by patenting. Finally, the firms compete in the product market by choosing prices.

2.3 Patent and antitrust policy

Due to the assumption of free entry and the existence of a large number of potential second-generation innovators, we assume that first- and second-generation innovators cannot make any agreements before the development of the second innovation.[10] After both innovations are developed, the government can affect the division of profits between different generations of innovators through its choice of policy. Assuming that the government can commit to a particular policy, the expected profits of firms depends on the specifics of that policy, namely how broad patent protection is and whether collusive licensing agreements are allowed.

Hence, in order to determine the optimal policy, we compare the following three policy regimes. Under a policy of infringement (I), if a second firm develops an improvement on a patented innovation and patents it, the courts find an infringement between the two sequential innovations and the two firms end up holding "blocking patents." Neither the first nor the second innovator can use the improved version without the other's permission. In such cases, the firms are allowed to maximize their joint profits. This is because when the courts find an infringement, the firms can achieve the collusive outcome without using any price restrictions. If one of the innovators licenses its innovation to the other innovator, the licensee can sell the improved version of the product as a monopolist.[11] This policy regime provides the first

[10] Such ex-ante agreements would allow them to share both the expected cost and benefit of the second innovation (Green and Scotchmer, 1995).

[11] The purpose of patent protection is to give the patentholder the right to exclude others from the use of the innovation covered by the patent. Legally, if the court decides that there is infringement, both the original innovation and the improvement are covered by the same patent and the courts maintain that the patentholder has the right to exclude others from the use of both of the products. We thank Howard Chang

innovator with the broadest patent protection.

If the courts do not find infringement (NI), we assume that they can further decide whether to allow the firms to enter a collusive agreement. Such collusive agreements may increase the investment incentives of innovators by eliminating the competition between holders of competing but non-infringing patents. Thus, we may have either a policy of no infringement and collusion (NIC), or a policy of no infringement and no collusion ($NINC$).[12,13]

Collusive agreements between firms can take many forms. The innovators can jointly maximize their profits by forming a patent pool, making a cross-licensing agreement, or simply giving one of the parties exclusive licenses to use the technologies owned by the innovators. The "Antitrust Guidelines for the Licensing of Intellectual Property" issued jointly by the Department of Justice (DOJ) and the Federal Trade Commission (FTC) describe generally how the agencies will handle various forms of licensing deals that may contain anticompetitive provisions. As pointed out in the Guidelines, such anticompetitive provisions are especially worrisome when the technology transfer takes place between horizontal competitors.[14] The general approach stated in the Guidelines is that such provisions will not be challenged as long as they are "necessary to achieve procompetitive benefits that outweigh... [the] anticompetitive effects (p. 12)." In practice, the approach has generally been to allow cross-licensing agreements and patent pools that include complementary patents, but to prohibit

for clarifying this point for us.

[12] Thus, we assume that the government adopts one of three possible policy regimes and that the firms know with certainty whether the courts will find infringement and whether they will allow collusive agreements. A more realistic representation of the government's intellectual property policy would be to allow for a continuum of cases and assume that the courts find infringement with probability α and allow collusive agreements with probability β. A similar kind of policy analysis can then be carried out by solving for the socially optimal α^* and β^*.

[13] In the following analysis, we assume that the patent system works efficiently and that the courts do not make mistakes in their interpretations of the law.

[14] Specifically, the Guidelines state the following. "Antitrust concerns may arise when a licensing arrangement harms competition among entities that would have been actual or likely potential competitors in a relevant market in the absence of the license (entities in a 'horizontal relationship'). A restraint in a licensing arrangement may harm such competition, for example, if it facilitates market division or price-fixing."

those that include substitute patents (Shapiro, 2001).

3 The Innovation Process

We are interested in finding the subgame perfect Nash equilibria of the game outlined in Section 2.2. Therefore, we start from the second race and work backwards. After analyzing the investment incentives of the firms under patenting and no patenting respectively, we determine the conditions under which patenting will take place. Finally, we move on to analyze the investment incentives in the first race.

3.1 Second race

3.1.1 Patenting

Suppose that the winner of the first race has patented and commercialized the first innovation. Let π^C stand for the interim profit it makes per period until the second race ends, which is equal to the value of the first innovation, v. Let Π^W and Π^L represent the discounted value of future profits from winning and losing the second race respectively. The first innovator's expected payoff function at the beginning of the second race is

$$\begin{aligned} V_w &= \int_0^\infty e^{-(\lambda X_2+r)t}\left(\lambda x_w \Pi^W + \lambda X_o \Pi^L + \pi^C\right) dt - c x_w \\ &= \frac{\lambda x_w \Pi^W + \lambda X_o \Pi^L + \pi^C}{\lambda X_2 + r} - c x_w \end{aligned} \tag{1}$$

where $X_o = \sum_i x_i$ represents the aggregate investment of all other firms, which are assumed to be symmetric, and $X_2 = X_o + x_w$ represents the aggregate investment in the second race. Since the research efforts of different firms are assumed to be independent of each other, we can simply add the individual investment levels to get the aggregate investment level.

If winner of the first race also wins the second race, it earns $2v$ per period. Thus, $\Pi^W = 2v/r$. The magnitude of Π^L depends on the policy regime. If the policy regime is I, the first innovator can make a positive profit even if it loses the second race. The second innovator has exclusive rights to the second innovation, but it cannot use it without making a licensing agreement with the first innovator. If the firms can reach an agreement, the second innovation can be sold at a price of $2v$, which is the maximum amount the consumers are willing to pay for the product. This is because if there is infringement, the two innovators are legally allowed to maximize joint profits. If the firms cannot reach an agreement, the first innovator will continue to sell the first innovation at a price of v. We assume the parties have equal bargaining power and divide the surplus to be shared, v, equally. Therefore, if the first innovator loses the second race, it can earn $\frac{v}{2}$ per period in addition to what it can make on its own without reaching an agreement with the second innovator.

Let $V_w^{P/I}$ denote the expected profits of the winner of the first race when it patents and the courts find an infringement. After substituting for the relevant values of Π^W, Π^L and π^C in (1), we get

$$V_w^{P/I} = \frac{\lambda x_w \left(\frac{2v}{r}\right) + \lambda X_o \left(\frac{v}{r} + \frac{v}{2r}\right) + v}{\lambda X_2 + r} - cx_w. \tag{2}$$

The innovator maximizes $V_w^{P/I}$ with respect to x_w. The first order condition gives[15]

$$\frac{\lambda v \left(\lambda X_o + 2r\right)}{2cr} = \left(\lambda X_2 + r\right)^2. \tag{3}$$

If one of the rival firms wins the second race, it earns $\frac{v}{2}$ per period. Its payoff function is

$$V_i^{P/I} = \frac{\lambda x_i \left(\frac{v}{2r}\right)}{\lambda X_2 + r} - cx_i. \tag{4}$$

Under symmetry, each rival firm maximizes $V_i^{P/I}$ with respect to x_i. Since there is free entry into the R&D race, the equilibrium individual investment levels and number of firms can be determined by solving the first order condition of the innovator, the first order condition of a generic rival firm, and the zero profit condition simultaneously. Since we are interested in

[15]It is straightforward to check that the second order condition is met.

the aggregate investment of all rival firms, we can use the zero profit condition to determine the rival firms' aggregate best response to the first innovator's investment choice. Combining the zero profit condition together with (3) yields $x_w^{P/I} = r/\lambda$ and $X_o^{P/I} = \left(\lambda v - 4cr^2\right)/2cr\lambda$.

Under policy regime NIC, the courts do not find an infringement but still allow the two winners, which hold competing patents, to collude in the product market. If the firms do not collude, price competition between the firms drives the price of the first innovation to zero and the price of the second innovation to v minus some infinitesimal amount. If the firms collude, they can charge $2v$ instead of v for the improved version. This implies that if a rival firm wins the second race, the first innovator receives $\frac{v}{2}$ and the second innovator receives $v + \frac{v}{2}$ per period. Thus, the second innovator receives half of the market value of the first innovation in addition to the marginal value of the second innovation. The first innovator, on the other hand, receives less than it receives under policy regime I if it loses the second R&D race.

The first innovator solves

$$\max_{x_w} \quad V_w^{P/NIC} = \frac{\lambda x_w \left(\frac{2v}{r}\right) + \lambda X_o \left(\frac{v}{2r}\right) + v}{\lambda X_2 + r} - cx_w \tag{5}$$

and a generic rival firm i solves

$$\max_{x_i} \quad V_i^{P/NIC} = \frac{\lambda x_i \left(\frac{v}{r} + \frac{v}{2r}\right)}{\lambda X_2 + r} - cx_i \tag{6}$$

respectively. Again, due to free entry, we can set $V_i^{P/NIC} = 0$ and solve for X_o. Substituting for X_o in the first derivative of the first innovator's payoff, we find that it is always negative. Therefore, the first innovator finds it optimal not to invest in the second race. We have $x_w^{P/NIC} = 0$ and $X_o^{P/NIC} = \left(3\lambda v - 2cr^2\right)/2cr\lambda = X_2^{P/NIC}$.

Finally, under policy regime $NINC$, the courts do not find an infringement and they do not allow the firms to collude in the product market. The first innovator does not earn anything if it loses the second race. The winner of the second race, if different from the winner of the first race, receives v per period, the marginal value of the innovation. This is higher than it earns under I, but lower than it earns under NIC since the firms are not allowed to collude. Solving the zero profit condition and the first innovator's first order condition simultaneously, we get $x_w^{P/NINC} = X_o^{P/NINC} = \left(\lambda v - cr^2\right)/2cr\lambda$.

Comparing the investment levels, we have $x_w^{P/NINC} > x_w^{P/I} > x_w^{P/NIC}$ and $X_o^{P/NIC} > X_o^{P/NINC} > X_o^{P/I}$. As shown in Table 1, for rival firms, the returns to investment are highest under NIC and lowest under I. If the courts rule that there is no infringement, the winner of the second race does not have to share its profits from the second innovation with the first innovator. Thus, the second innovator has a higher bargaining power in case of no infringement. If, after finding no infringement, the courts also allow collusive licensing agreements, the second innovator gets half of the surplus that the consumers would enjoy otherwise, which is equal to half of the value of the first innovation. Thus, rival firms' incentives to invest are higher under NIC than they are under $NINC$.

	Payoff to first innovator	Payoff to firm i
I	$\frac{v}{r} + \frac{v}{2r}$	$\frac{v}{2r}$
$NINC$	0	$\frac{v}{r}$
NIC	$\frac{v}{2r}$	$\frac{v}{r} + \frac{v}{2r}$

Table 1: Payoffs if firm i wins the second R&D race

The first innovator, while choosing its investment level, also takes into account its expected profits from losing in the second race, which depends on the intensity of the competition it faces. It invests more aggressively if its expected profits from losing are low. Thus, the incentives to win are highest under $NINC$ since it gets zero if it loses. As far as I and NIC are concerned, we can see from Table 1 that it can collect a higher payoff under I than under NIC if it loses in the second race. However, its *expected* payoff is higher under NIC than it is under I because rival firms have higher incentives to invest under NIC. Thus, it prefers not to invest at all under NIC while it still invests a positive amount under I.

These results have an important implication for the dynamic evolution of the market.[16] Comparing the first innovator's and rival firms' investment levels yields that $X_o^{P/NIC} > X_o^{P/NINC} = x_w^{P/NINC} > x_w^{P/NIC}$. Allowing collusion in case of no infringement substantially decreases the first innovator's incentive to invest (basically to zero) and increases rival firms' incentives to invest in the second race. This implies that while the first innovator has the same chance of winning the second race as rival firms under $NINC$, the identity of the market leader will certainly change under NIC. Thus, whether the identity of the market leader changes over time depends on the nature of intellectual property policy.

The investment levels in the subgame where the first innovation has been patented but not commercialized can be found in a similar way. If the first innovation is not commercialized, then $\pi^C = 0$. As far as the commercialization decision is concerned, we need to compare the profits under commercialization with those under no commercialization. Straightforward analysis shows that if the winner of the first race patents the first innovation, it always commercializes it and starts to earn interim profits.

3.1.2 No patenting

The first-to-invent rule implies that in case of an independent re-invention, the first innovator can still claim exclusive ownership over the first innovation as long as it can prove its first-to-invent status. Since it has nothing to lose by patenting, the first innovator prefers to patent the first innovation if the second innovation is developed by a rival. Thus, its payoffs are still as given in Table 1.

[16] I am grateful to an anonymous referee for emphasizing this implication of the results. The issue of persistence of monopoly was first addressed in Gilbert and Newbery (1982) and Reinganum (1983).

If the first innovator has not patented or commercialized the first innovation, (1) becomes

$$V_w = \frac{\lambda x_w \Pi^W + \mu X_o \Pi^L}{\lambda x_w + \mu X_o + r} - c x_w. \tag{7}$$

This expected payoff function differs from (1) in two ways. First, if the first innovation is not patented, the rival firms' hazard rate decreases. Second, the first innovator cannot make any interim profits from the first innovation until the second innovation is developed.

Commercialization results in reverse engineering. This implies that the first innovator's interim profit is zero and rival firms' hazard rate is λX_o instead of μX_o. Moreover, since the first innovator will not be able to patent the first innovation in case of independent re-invention, Π^L is equal to zero. Thus, its expected payoff is

$$V_w = \frac{\lambda x_w \Pi^W}{\lambda x_w + \lambda X_o + r} - c x_w. \tag{8}$$

The investment levels under the three policy regimes can be found following the same line of analysis as in Section 3.1.1. Comparing the profit level under commercialization with the profit level under no commercialization, we discover that if the policy regime is $NINC$, the first innovator always chooses not to commercialize. This is because $\Pi^L = 0$ whether or not it patents. Not commercializing has the benefit of lowering the hazard rate of the rival firms, so it is strictly preferable. If the policy regime is I, the first innovator will choose not to commercialize if and only if $\left(3\lambda^2 - 2\lambda\mu + \mu^2\right) v - 6cr^2\lambda \geq 0$. If the policy regime is NIC, the first innovator will choose not to commercialize if and only if $9\left(\lambda - \mu\right)^2 v - 2cr^2\lambda \geq 0$. These inequalities show that not commercializing is more attractive if the first innovator can significantly reduce the competition it faces with secrecy, i.e., if μ is low.

3.2 Patenting decision

We know from Section 3.1.1 that if the first innovation is patented, profits under commercialization are higher than profits under no commercialization. Moreover, we can easily show that patenting and commercializing strictly dominates not patenting and commercializing under policy regimes I and NIC. This is because patenting under I and NIC allows the first innovator to have a positive payoff even if it loses the second race. If the policy regime is $NINC$, the two strategies are equivalent since $\Pi^L = 0$ under $NINC$.

Thus, the winner of the first race decides whether to patent by comparing the profits from patenting and commercializing with the profits from not patenting and not commercializing. The benefit of patenting is that the innovator can start to make profits from the innovation immediately. The benefit of delaying patenting is that it slows down the progress of rival firms. At the end of the first R&D race, the winner decides whether to patent by considering this trade-off. Clearly, μ, the rival firms' marginal benefit of investment in case of no patenting, plays a crucial role in the patenting decision.

Lemma 1 π_w^{NP} *is a decreasing function of* μ.

Under all of the policy regimes, the profits of the first innovator under no patenting decreases as μ increases and competition in the R&D market increases. If the first innovator can sufficiently impair the positions of the rival firms in the second R&D race by not patenting, it will choose not to patent. Let μ^I stand for the critical μ value such that for $\mu > \mu^I$, the first innovator will choose to patent under regime I. We can define μ^{NIC} and μ^{NINC} in a similar fashion. Proposition 1 states how the incentives to patent differ among the three policy regimes.

Proposition 1 *The first innovator's threshold for patenting is lowest under regime I and highest under regime NINC:* $\mu^I < \mu^{NIC} < \mu^{NINC}$.

Providing first-generation innovators with strong protection increases their incentives to patent. The regimes differ in the amount of licensing fees patentholders can collect if a rival firm develops an improved version of their innovation. These fees are highest under I and equal to zero under $NINC$. Not patenting decreases the investment incentives of the rival firms and relaxes the competition in the R&D industry. The first innovator is less concerned about the intensity of competition in the R&D market the higher the amount it will earn if it loses the second race. Hence, its incentive to patent is highest under policy regime I because it provides the innovator with the highest insurance against losing in the second R&D race.

Thus, we are likely to see patenting more frequently under regime I. There exists values of μ for which the first innovator chooses to patent under regime I while it chooses not to patent under the other two policy regimes.

Comparing the investment levels under patenting and commercializing with those under no patenting and no commercializing, we observe that under all policy regimes, rival firms have lower incentives to invest if the first innovation is not patented. This is because of the decrease in the marginal benefit of investment under secrecy ($\mu \leq \lambda$). For the winner of the first race, not patenting affects its investment incentive in three different ways. First, it faces a lower level of competition in the R&D market. Second, it cannot make any interim profits from the first innovation, so it is more impatient to bring the second race to an end. Third, since rival firms have decreased incentives to invest, the first innovator's expected payoff in the case of losing in the second race is lower.

If the policy regime is I or NIC, the second and third effects cause x_w to increase. The role of the first effect is ambiguous.[17] Comparing the investment levels shows that under policy regimes I and NIC, the first innovator always invests more if it does not patent. Hence, the second and third effects dominate the first effect in cases when they work in opposite directions.

[17] To see this, consider the slope of the reaction function of the first innovator. Using the first order condition, we can show that sign $\{\partial x_w / \partial X_o\} = $ sign $\left\{(\lambda x_w - \mu X_o)\left(\pi^W - \pi^L\right) - r\left(\pi^W + \pi^L\right)\right\}$. Thus, the investment levels are likely to be strategic complements for low levels of X_o and strategic substitutes for high levels of X_o.

If the policy regime is $NINC$, the first innovator invests more under no patenting if $(\lambda-\mu)^2 v - cr^2\lambda > 0$. That is, the first innovator invests more under no patenting for μ close to λ and invests less under no patenting for low values of μ. This is because when the policy regime is $NINC$, the first innovator earns zero if it loses the second race whether or not it patents. Thus, the third effect stated above is no longer present. Moreover, the first effect is weak when μ is high since not patenting does not have a significant effect on the amount of competition the first innovator faces in the R&D stage. Thus, it invests more under no patenting mainly due to the second effect. When μ is low, the first and second effects work in opposite directions since the investment levels are strategic complements. The first effect dominates and the first innovator invests less under no patenting. In other words, the first innovator can considerably decrease the amount of competition it faces in the R&D stage by not patenting. Therefore, it does not have to invest as much under no patenting as it does under patenting.

In the following analysis, we focus on the case where the aggregate investment of the rival firms under both patenting and no patenting is always positive. Since their incentives to invest are lowest under regime I, this implies that $\lambda v > 4cr^2$ in case of patenting and $\mu^2 v > 8cr^2\lambda$ in case of no patenting.

3.3 First race

We can now analyze the investment incentives in the first R&D race. If a firm loses the first race, it can still participate in the second R&D race. However, its expected profit in the second race is zero due to free entry.

Since all of the firms are symmetric in the first race, each firm has a payoff function of the form

$$V_i = \frac{\lambda x_i V_w}{\lambda X_1 + r} - cx_i \tag{9}$$

where X_1 represents the aggregate investment level in the first race. The magnitude of V_w, the winner's expected profit at the beginning of the second race, depends on whether the innovation is patented and what the policy regime is.

Proposition 2 compares the aggregate investment levels under the three policy regimes in the first as well as the second R&D race.

Proposition 2 *(i) In the first R&D race,* $X_1^{P/I} > X_1^{P/NINC} > X_1^{P/NIC}$ *and* $X_1^{NP/I} > X_1^{NP/NINC} > X_1^{NP/NIC}$*. (ii) In the second R&D race,* $X_2^{P/NIC} > X_2^{P/NINC} > X_2^{P/I}$ *and* $X_2^{NP/NIC} > X_2^{NP/NINC} > X_2^{NP/I}$*.*

While regime NIC favors second-generation investors at the expense of first-generation investors, regime I favors first-generation investors at the expense of second-generation investors. To see this, note that due to free entry, the aggregate investment level in the first race can be determined by using the zero profit condition. We get

$$X_1 = \frac{V_w}{c} - \frac{r}{\lambda} \tag{10}$$

which implies that the aggregate investment level in the first race is increasing in the first innovator's expected profit level in the second race. This is why the investment incentives are maximized under regime I. Comparing the expected profit of the winner under regime NIC and $NINC$, we observe that $V_w^{P/NINC} > V_w^{P/NIC}$ and $V_w^{NP/NINC} > V_w^{NP/NIC}$. Since the winner of the first race faces stronger competition in the second race under NIC, its expected profits are lower in that case. This causes the investment level in the first race to be lower under regime NIC.

4 The Optimal Policy

This section examines optimal policy by focusing on expected social welfare at the beginning of the first R&D race. Social welfare is defined as the sum of producer surplus, consumer surplus, and the non-appropriable values of the innovations. As pointed out in Arrow (1962), for a variety of reasons investors may not always be able to appropriate for themselves the entire social benefit of their innovations. Let $s \geq 0$ stand for the non-appropriable value of the innovations. It represents the increase in social welfare that firms in other industries and their consumers may enjoy due to either knowledge or demand spillovers.[18]

In writing the expected social welfare function, it is necessary to take into account the spillovers between the first and the second innovations. As pointed out by Green and Scotchmer (1995) and Chang (1995), the social benefit from an innovation should include the option value of the subsequent innovations it inspires since those innovations would not have been possible without the first one. Thus, if the first innovation is patented, expected social welfare is

$$W = P\left(X_1\right)\left[\left(\frac{v+s}{r}\right) + \left(P\left(X_2\right)\left(\frac{v+s}{r}\right) - cX_2\right)\right] - cX_1 \tag{11}$$

where $P\left(X_1\right) = \lambda X_1 / \left(\lambda X_1 + r\right)$ and $P\left(X_2\right) = \lambda X_2 / \left(\lambda X_2 + r\right)$. As in Denicolo (2000), these expressions can be interpreted as the probability of innovating adjusted by the discount rate.

Due to free entry, expected profits are bid down to zero in both of the races. Social welfare becomes a function of the non-appropriable values of the innovations and consumer surplus. We have

$$\begin{aligned} W^{P/I} &= P\left(X_1^{P/I}\right)\left[\frac{s}{r} + P\left(X_2^{P/I}\right)\frac{s}{r}\right] \qquad (12)\\ W^{P/NIC} &= P\left(X_1^{P/NIC}\right)\left[\frac{s}{r} + P\left(X_2^{P/NIC}\right)\frac{s}{r}\right] \\ W^{P/NINC} &= P\left(X_1^{P/NINC}\right)\left[\frac{s}{r} + P\left(X_2^{P/NINC}\right)\frac{s}{r} + \left(\frac{\lambda X_o^{P/NINC}}{\lambda X_2^{P/NINC} + r}\right)\frac{v}{r}\right]. \end{aligned}$$

The extra term in $W^{P/NINC}$ represents consumer surplus. If the firms are not allowed to collude, there is competition between the first and second innovators. Consumers receive a

[18] Jones and Williams (1998) find that the optimal R&D investment is at least two to four times actual investment.

surplus of v per period after the second innovation is developed. Consumer surplus exists if and only if the second innovation is developed by one of the rival firms. This is why we have $\lambda X_o^{P/NINC}$ instead of $\lambda X_2^{P/NINC}$ in the numerator. Under policy regimes I and NIC, the investors can capture all of the consumer surplus, which increases their willingness to invest. Since consumers are assumed to buy at most one unit of the good, charging a price above marginal cost does not result in any deadweight loss.

If the first innovation is not patented, we have

$$\begin{aligned} W^{NP/I} &= P\left(X_1^{NP/I}\right) P\left(X_2^{NP/I}\right) \frac{2s}{r} \\ W^{NP/NIC} &= P\left(X_1^{NP/NIC}\right) P\left(X_2^{NP/NIC}\right) \frac{2s}{r} \\ W^{NP/NINC} &= P\left(X_1^{NP/NINC}\right) \\ & \left[P\left(X_2^{NP/NINC}\right) \frac{2s}{r} + \left(\frac{\mu X_o^{NP/NINC}}{\lambda x_w^{NP/NINC} + \mu X_o^{NP/NINC} + r}\right) \frac{v}{r}\right]. \end{aligned} \tag{13}$$

Under no patenting, since the first innovator does not use the innovation for commercial purposes until the second innovation is developed, the social benefit of the first innovation as well as the second innovation can be realized only after the second innovation is developed. This is the main difference between the welfare functions under patenting and no patenting.

We first analyze whether the first innovator's patenting decision is socially optimal. Since in R&D race models the private incentives to invest may exceed the social incentives to invest, disclosure of innovations, which increases the investment incentives of innovators, may not always be desirable. Proposition 3 states that in our model, whenever the first innovator decides to patent, it makes the socially optimal choice.

Proposition 3 *Under all of the policy regimes, if* $V_w^P > V_w^{NP}$*, then* $W^P > W^{NP}$*.*

In cases when the first innovator prefers patenting to secrecy, the investment incentives of both first-generation and second-generation innovators are higher under patenting than under secrecy. This makes patenting socially desirable since the social returns from innovation exceed the private returns from innovation and under free entry, social welfare is a function of the non-appropriable values of the innovations only.

We next compare the relevant welfare functions under the different μ values in order to determine the optimal intellectual property policy.

Proposition 4 *(i) For $\mu \leq \mu^I$, $W^{NP/NIC} > W^{NP/I}$. Moreover, there exists $s^{low} > 0$ such that $W^{NP/NIC} > W^{NP/NINC}$ for $s > s^{low}$. (ii) For $\mu^{NINC} \geq \mu > \mu^I$, $W^{P/I} > W^{NP/NIC}$ (if $\mu \leq \mu^{NIC}$) and $W^{P/I} > W^{P/NIC}$ (if $\mu > \mu^{NIC}$). Moreover, there exists $s^{int} > 0$ such that $W^{P/I} > W^{NP/NINC}$ for $s > s^{int}$. (iii) For $\mu > \mu^{NINC}$, $W^{P/I} > W^{P/NIC}$. Moreover, there exists $s^{high} > 0$ such that $W^{P/I} > W^{P/NINC}$ for $s > s^{high}$.*

Table 2 summarizes the results presented in Proposition 4 together with the results presented in Proposition 1.

	$\mu \leq \mu^I$	$\mu^I < \mu \leq \mu^{NIC}$	$\mu^{NIC} < \mu \leq \mu^{NINC}$	$\mu^{NINC} < \mu \leq \lambda$
Patenting under:	none	I only	I and NIC only	I, NIC, and $NINC$
Optimal policy:	$s \leq s^{low} : NINC$ $s > s^{low} : NIC$	$s \leq s^{int} : NINC$ $s > s^{int} : I$	$s \leq s^{int} : NINC$ $s > s^{int} : I$	$s \leq s^{high} : NINC$ $s > s^{high} : I$

Table 2: Patenting incentives and optimal policy under the three regimes

There are several things to stress about Proposition 4. Consider first optimal policy when the non-appropriable value of the innovations, s, is sufficiently high. Proposition 4 states that it is optimal to have narrow patent protection and collusion (NIC) for $\mu \leq \mu^I$ and broad patent protection (I) for $\mu > \mu^I$. In cases when s is sufficiently high, the benefit of having faster development of the innovations outweighs the harm caused by the reduction in the consumer surplus. Hence, it is socially desirable to transfer rents from consumers to producers in order to encourage investment in the two innovations. This can be done by having either policy regime NIC or I. Policy regime I favors first-generation innovators while policy regime NIC favors second-generation innovators. It is more desirable to have NIC when the first innovator prefers secrecy under all the policy regimes. This implies that it is not necessary to protect the interests of the first innovator when the first innovator can protect itself by secrecy.[19] It is desirable to protect the interests of the second-generation investors instead since non-disclosure causes them to have substantially lower incentives to invest. The aggregate incentives to invest in the second R&D race are highest under policy regime NIC since it allows the second innovator to receive a share of the value of the first innovation. This result differs from those in the literature, which are mainly concerned about how to allocate the value of the second innovation between the two innovators.[20]

For $\mu > \mu^I$, there will always be patenting under policy regime I. From Table 2 we can see that for this range of μ, welfare under I dominates welfare under NIC whether or not there will be patenting under NIC. Since the development of the first innovation motivates the development of the second innovation, when the values of the innovations are symmetric, it is desirable to have a policy regime which gives more weight to the development of the first innovation. In such cases, policy regime NIC provides second-generation innovators with unnecessarily high incentives to invest.[21]

[19] It is easy to check that for $\mu \leq \mu^I$, the ranking between I and NIC does not depend on the assumption of symmetry between the non-appropriable values of innovations. That is, no matter how large s_1 is, it is never optimal to have I if there will not be patenting under I.

[20] In cases when μ is so low that the innovator can deter entry into the second R&D race by not patenting, there may be more scope for allowing collusive licensing agreements because the rivals would have more incentives to enter the second race under regime NIC than they would under the other two regimes.

[21] One question that arises is whether there are regions of the parameter space where policy regime NIC is desirable if we relax the symmetry assumption. Given the trade-off between the policy regimes I and NIC, one would expect the ratios $\frac{W^{P/I}}{W^{P/NIC}}$ and $\frac{W^{P/I}}{W^{NP/NIC}}$ to decrease as $\frac{s_2}{s_1}$ increases. In fact, it is straightforward to show that *ceteris paribus*, this is exactly the case. As the non-appropriable value of the second innovation gets relatively higher, policy regime NIC is more likely to be socially desirable.

When the non-appropriable value of the innovations, s, is sufficiently low, it is optimal to have $NINC$, which allows consumers to benefit from the competition between sequential innovators. To analyze how optimal policy changes as s increases, consider the region where $\mu \leq \mu^I$. Under no patenting, society cannot start to benefit from the innovations until both of them are achieved. Therefore, the goal is to have the second innovation developed as soon as possible. Proposition 2 states that, compared with policy regime $NINC$, under policy regime NIC the first-generation innovators have lower incentives to invest while the second-generation innovators have higher incentives to invest. Thus, regime NIC becomes more desirable as s increases.

Now consider the region where $\mu > \mu^I$. For $\mu^I < \mu \leq \mu^{NINC}$, there will be patenting under I but not under $NINC$. Since under policy regime I society can start to benefit from the first innovation without any delay, it becomes socially desirable to have I as the non-appropriable value of the innovations increases. For $\mu^{NINC} < \mu \leq \lambda$, there will be patenting under both I and $NINC$. However, as s increases, it again becomes socially desirable to have regime I because first-generation innovators have higher incentives to invest while second-generation innovators have lower incentives to invest under policy regime I. This implies that society is expected to start to benefit from the first innovation sooner under I. Moreover, faster development of the first innovation means a faster start for the development of the second innovation.

Having established the desirability of regime I for $\mu > \mu^I$ and sufficiently high values of s, we next explore whether the desirability of I is sensitive to the magnitude of μ. Does regime I become less attractive as μ increases and disclosure through patenting becomes more likely? The size of μ depends on the government's trade secret policy and the firms' ability to monitor the flow of their technological information. Clearly, different industries may be characterized by different values of μ because the effectiveness of secrecy as a protection mechanism may vary across industries. Proposition 5 implies that having strong patent protection becomes less desirable as μ increases.

Proposition 5 *For all values of μ, $s^{high} > s^{int}$.*

If the probability of imitation under secrecy is high, the innovator has an increased incentive to patent the innovation in order to protect itself. In fact, when μ is sufficiently high, there will be patenting under all of the policy regimes because the first innovator cannot significantly decrease the success rate of the rival firms by not patenting. Therefore, Proposition 5 illustrates that if early innovations are likely to be disclosed through patenting whether or not there is broad patent protection, having narrow patent protection becomes more attractive from a social point of view. Having broad patent protection may not be as desirable because it stifles the investment incentives of second-generation innovators. However, if the probability of imitation under secrecy is low and early innovations are likely to be kept secret in the absence of broad patent protection, having broad patent protection may actually increase social welfare by encouraging earlier disclosure of innovations.

In the next section we review the related literature and discuss how our results compare with the existing results in the literature.

5 Comparison with Earlier Literature

One of the main issues raised in the literature on cumulative innovation has been that if future innovations are improved versions of earlier innovations, the incentives to develop the earlier innovations may be too low. A natural solution is to have broad patent protection (Kitch, 1977). However, such broad protection may have the effect of stifling the investment incentives of subsequent innovators (Scotchmer, 1991). Thus, Merges and Nelson (1990) suggest that because of the hold-up problem, the courts should find infringement when the original innovation has a large stand-alone value only. Green and Scotchmer (1995) point out that the hold-up problem can be overcome if sequential innovators are allowed to have ex-ante licensing contracts after the first innovation is developed but before the second innovator invests. Since such ex-ante contracts raise antitrust issues and may not always be feasible, Chang (1995) analyzes optimal patent policy assuming ex-ante licensing is not possible. He shows that the courts' infringement standards should extend the broadest protection to innovations with very large (similar to Merges and Nelson, 1990) and very small stand-alone values relative to the improvements that they may inspire.

All these papers assume that each innovation is developed by a different firm and that there is no racing in the development of the innovations.[22] Denicolo (2000) analyzes the optimal patent breadth in a model that allows for competition in the R&D stage and repeated innovation by the same firm.[23] He shows that broad patent protection becomes less attractive as the relative profitability of the first innovation increases since the firms then have stronger incentives to invest.

This paper extends the discussion on optimal patent breadth by pointing out that the desirability of broad patent protection also depends on the patenting incentives of early innovators. In the previous literature, broad patent protection is seen as a way of stimulating investment in early innovations. Broad patent protection can also be used to encourage patenting of early innovations. Thus, we have shown that as the effectiveness of secrecy as a protection mechanism increases, i.e., as μ decreases, it is optimal to have broad patent protection over a larger set of parameters. Our results complement those of Denicolo (2000)

[22] This may be a realistic assumption to make especially if the second-generation product is an application of the first innovation. However, in many cases in the real world, where the second-generation product is an improvement of the first innovation, it may not be realistic to assume that initial innovators do not invest to improve their products. The main distinction between an application and an improvement is that an application is assumed not to affect the profits made from the original innovation while an improvement is assumed to reduce the profits made from the original innovation.

[23] Two other exceptions are Scotchmer and Green (1990), and O'Donoghue et al. (1998). However, they focus on the analysis of different policy instruments.

by analyzing optimal policy in industries where early innovations may not be patented.

If early innovations are not likely to be patented under any of the policy regimes, we have shown that having narrow patent protection and allowing collusion may be optimal. As far as antitrust policy is concerned, it is informative to compare our results with those of Chang (1995) since both papers consider the same policy regimes. There are three main differences between the current paper and Chang (1995). First, Chang (1995) notes that for the first innovator, regime $NINC$ is dominated by all others while in our case, the investment incentives of first-generation innovators are the lowest under NIC. This is due to our assumption of competition in the R&D stage. The first innovator will face more intense competition in the second R&D race and, thus, has lower expected profits from winning in the first R&D race under NIC compared to $NINC$. Second, in Chang (1995) the second innovator has an excessive incentive to innovate under NIC (because it receives a share of the value of the first innovation) while in our model this is not necessarily the case. There are two reasons for this discrepancy. First, we assume that the first innovation may not always be patented. Thus, when the first innovator chooses secrecy and reduces the investment incentives of subsequent innovators, having regime NIC may actually be desirable in order to stimulate investment in the second innovation. The second reason is that the social value of an innovation in our paper is not equal to its private value whereas in Chang (1995) it is. Since the gap between the social and private returns from innovation (s) may result in underinvestment, having regime NIC becomes more desirable as this gap increases.

The third difference between our results and those of Chang (1995) is that he argues that collusion should be allowed in order to help first-generation innovators while we argue that it should be allowed in order to help second-generation innovators.[24] In his model, policy regime NIC, compared with policy regime $NINC$, affects the profits of the first innovator

[24] Scotchmer (1991) also suggests that in case of competing innovations, antitrust authorities may find it optimal to allow collusive licensing agreements in order to encourage investment in the first innovation.

in two contradictory ways. On the one hand, the second innovator has an excessive incentive to invest under NIC, which reduces the profits of the first innovator. On the other hand, the first innovator can still receive a share of the value of the first innovation after the second innovator develops the second innovation under NIC.[25] If the value of the first innovation is low, the second innovator's excessive incentive to invest is not very high and the second effect dominates the first. Hence, Chang (1995) concludes that collusion is desirable, to help the first innovator, if the (appropriable) value of the first innovation is very small. However, we find that collusion is desirable, to help second-generation investors, if secrecy is chosen under all policy regimes considered and the (non-appropriable) value of the second innovation is sufficiently high.[26]

In a recent paper, Denicolo (2002) also finds that collusion may be desirable if the non-appropriable value of the second innovation is high. However, he does not take into account the patenting incentive of the first innovator, and assumes that the first innovator and the rival firms invest sequentially. Finally, Priest (1977) and Kaplow (1984) give no support for collusive licensing deals on the grounds that they reward innovation at an excessive social cost. However, neither paper considers the cumulative nature of innovation.

6 Concluding Remarks

This paper has examined optimal policy assuming innovators can try to protect their innovations by relying on either the patent system or secrecy. Although the patent system provides innovators with the right to exclude others from using their innovations for a fixed period of time, it also requires them to disclose their innovations. The literature on cumulative innovation has so far explored how broad patent protection and collusive agreements can be used to

[25] Note that in Chang (1995), the idea for each innovation occurs exogenously to only one firm.

[26] Chang (1995) concludes that since collusive licensing agreements are desirable only in a very limited number of cases, "a uniform rule against collusion would probably be the best policy (p. 49)."

give early innovators sufficient incentives to invest without stifling investment in subsequent innovations. The main contribution of this paper is to emphasize that in cases when early innovators prefer to protect their innovations by secrecy, the design of optimal patent and antitrust policy should pay attention to encouraging disclosure and stimulating investment in subsequent innovations.[27]

Specifically, we have shown that stronger patent protection may be preferable in industries where secrecy is an effective protection mechanism in order to encourage disclosure of earlier innovations. Moreover, we have demonstrated how antitrust policy can complement patent policy in stimulating innovation in industries where patent policy is ineffective in achieving disclosure. This suggests that having a lenient antitrust policy may be desirable not only in case of complementary patents, but also in case of substitute patents. Our results in general imply that policy geared towards industries where innovators can rely on secrecy to protect their innovations should be different from policy geared towards industries where innovators cannot rely on secrecy.

Our analysis can be extended in several ways by relaxing some of the assumptions we have made. First, we have assumed that the first-to-invent rule determines ownership. Currently, there is a lot of debate on whether the United States should continue with its first-to-invent policy or adapt the first-to-file policy to be in harmony with the rest of the world. Therefore, we can analyze what a switch to the first-to-file rule would imply in terms of optimal patent policy. The first-to-invent rule gives innovators more flexibility about when to disclose their innovations. Under the first-to-file rule, innovators may feel more threatened and have higher incentives to patent. Therefore, having strong patent protection may become less attractive. Second, we have assumed that reverse engineering is easy. If reverse engineering is difficult,

[27] Other papers that take into consideration the patenting incentives of innovators are Denicolo and Franzoni (2002), Matutes et al. (1996), Scotchmer and Green (1990), Gallini (1992), and Horstmann et al. (1985). However, the policy issues addressed in these papers differ significantly from those analyzed in the current paper.

the innovator would have a lower incentive to patent and the courts may find it optimal to have strong patent protection more frequently. Finally, we have analyzed optimal patent and antitrust policy taking the government's trade secret policy as given. The paper can be extended to analyze optimal trade secret policy. In such a more fully elaborated analysis, having a strong trade secret policy may be attractive because it prevents duplication of effort.

References

[1] Aoki, R. and Y. Spiegel. 2003. "Pre-Grant Patent Publication, R&D, and Welfare," available at http://www.tau.ac.il/~spiegel/papers/reiko.html.

[2] Arrow, K. J. 1962. "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, R. R. Nelson, ed., Princeton University Press.

[3] Arundel, A. 2001. "The Relative Effectiveness of Patents and Secrecy for Appropriation," *Research Policy*, 30, 611-624.

[4] Chang, H. F. 1995. "Patent Scope, Antitrust Policy, and Cumulative Innovation," *RAND Journal of Economics*, 26, 34-57.

[5] Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2000. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)," Working Paper No. W7552, National Bureau of Economic Research, 2000.

[6] Denicolo, V. 2000. "Two-Stage Patent Races and Patent Policy," *RAND Journal of Economics*, 31(3), 488-501.

[7] Denicolo, V. 2002. "Sequential Innovation and the Patent-Antitrust Conflict," *Oxford Economic Papers*, 54, 649-668.

[8] Denicolo, V. and P. Zanchettin. 2002. "How Should Forward Patent Protection be Provided?" *International Journal of Industrial Organization*, 20, 801-27.

[9] Denicolo, V. and L. A. Franzoni. 2002. "Patents, Secrets, and the First Inventor Defense," http://ssrn.com/abstract=265745.

[10] Gallini, N. T. 1992. "Patent Policy and Costly Imitation," *Rand Journal of Economics*, 23(1), 52-63.

[11] Gallini, N. T. 2002. "The Economics of Patents: Lessons from Recent U.S. Patent Reform," *Journal of Economic Perspectives*, 16(2), 131-154.

[12] Gilbert, R. J. and D. M. G. Newbery. 1982. "Preemptive Patenting and Persistence of Monopoly," *American Economic Review*, 72(3), 514-526.

[13] Green, J., and S. Scotchmer. 1995. "On the Division of Profit in Sequential Innovation," *RAND Journal of Economics*, 26, 20-23.

[14] Horstmann, I., G. M. MacDonald, and A. Slivinski. 1985. "Patents as Information Transfer Mechanisms: To Patent or (Maybe) Not to Patent," *Journal of Political Economy*, 93(5), 837-858.

[15] Jones, C. I. and J. C. Williams. 1998. "Measuring the Social Return to R&D," *Quarterly Journal of Economics*, 113, 1119-1135.

[16] Kaplow, L. 1984. "The Patent-Antitrust Intersection: A Reappraisal," *Harvard Law Review*, 97, 1813-1892.

[17] Kitch, E. W. "The Nature and Function of the Patent System," *Journal of Law and Economics*, 20, 265-290.

[18] Lerner, J. 1994. "The Importance of Trade Secrecy: Evidence from Civil Litigation," Harvard Business School Working Paper #95-043.

[19] Levin, R. C., A. K. Klevorick, R. R. Nelson, and S. G. Winter. 1987. "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3, 783-820.

[20] Loury, G. 1979. "Market Structure and Innovation," *Quarterly Journal of Economics*, 93, 395-410.

[21] Mansfield, E. 1985. "How Rapidly Does New Technology Leak Out," *The Journal of Industrial Economics*, 34, 217-223.

[22] Matutes, C., P. Regibeau, and K. E. Rockett. 1996. "Optimal Patent Protection and the Diffusion of Innovation," *Rand Journal of Economics*, 27, 60-83.

[23] Merges, R. P. and R. R. Nelson. 1990. "On the Complex Economics of Patent Scope," *Columbia Law Review*, 90(4), 839-917.

[24] O'Donoghue, T. 1998. "A Patentability Requirement for Sequential Innovation," *Rand Journal of Economics*, 29(4), 654-679.

[25] O'Donoghue, T., S. Scotchmer, and J.-F. Thisse. 1998. "Patent Breadth, Patent Life, and the Pace of Technological Progress," *Journal of Economics and Management Strategy*, 7(1), 1-32.

[26] Pitofsky, R. 2000. "Challenges of the New Economy: Issues at the Intersection of Antitrust and Intellectual Property," *Antitrust Law Journal*, 68, 913-924.

[27] Priest, G. L. 1977. "Cartels and Patent License Arrangements," *Journal of Law and Economics*, 20, 309-377.

[28] Reinganum, J. 1983. "Uncertain Innovation and the Persistence of Monopoly," *American Economic Review*, 73(4), 741-748.

[29] Scotchmer, S. and J. Green. 1990. "Novelty and Disclosure in Patent Law," *Rand Journal of Economics*, 21, 131-46.

[30] Scotchmer, S. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 5, 29-41.

[31] Scotchmer, S. 1996. "Protecting Early Innovators: Should Second-Generation Products be Patentable?" *Rand Journal of Economics*, 27, 322-31.

[32] Shapiro, C. 2001. "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting," in *Innovation Policy and the Economy*, Vol. 1, A. Jaffe, J. Lerner, and S. Stern, eds., MIT Press.

[33] Tom, W. K. 1998. "Licensing and Antitrust: Common Goals and Uncommon Problems," Address before the American Conference Institute, *9th National Conference on Licensing Intellectual Property*, http://www.ftc.gov/speeches/other/aciippub.htm.

[34] U.S. Department of Justice and Federal Trade Commission. 1995. "Antitrust Guidelines for the Licensing of Intellectual Property," http://www.usdoj.gov/atr/public/guidelines/ipguide/htm.

Appendix

A Proof of Lemma 1

Differentiating we get:

$$\begin{aligned}\frac{\partial V_w^{NP/I}}{\partial \mu} &= \frac{-(\lambda-\mu)v}{r\lambda^2}\\ \frac{\partial V_w^{NP/NIC}}{\partial \mu} &= \frac{-3(\lambda-\mu)v}{r\lambda^2}\\ \frac{\partial V_w^{NP/NINC}}{\partial \mu} &= \frac{-(2\lambda-\mu)v}{r\lambda^2}.\end{aligned}$$

Since $\lambda \geq \mu$ by assumption, the derivative is negative in all three cases.

B Proof of Proposition 1

Setting $V_w^P - V_w^{NP} = 0$ under the three policy regimes, we find

$$\begin{aligned}\mu^I &= \lambda - \frac{2r\sqrt{c\lambda v}}{v}\\ \mu^{NIC} &= \lambda - \frac{2r\sqrt{c\lambda v}}{3v}\\ \mu^{NINC} &= 2\lambda - \frac{\sqrt{\lambda v\left(\lambda v + cr^2\right)}}{v}.\end{aligned}$$

Clearly, $\mu^{NIC} > \mu^I$. Moreover,

$$\mu^{NINC} - \mu^{NIC} = \frac{\sqrt{\lambda}\left[3\sqrt{\lambda v} + 2\sqrt{cr^2} - 3\sqrt{\lambda v + cr^2}\right]}{3\sqrt{v}},$$

which is positive if $3\sqrt{\lambda v} + 2\sqrt{cr^2} > 3\sqrt{\lambda v + cr^2}$. Squaring both sides we get

$$9\lambda v + 4cr^2 + 12\sqrt{\lambda v cr^2} \overset{?}{>} 9\lambda v + 9cr^2.$$

Since $X_0^{P/I} > 0$ implies $\lambda v > 4cr^2$, this inequality always holds. Therefore, $\mu^{NINC} > \mu^{NIC}$. From Lemma 1 it follows that for all values of μ greater than these critical μ values, there will be patenting under the respective policy regimes.

C Proof of Proposition 2

(i) In the second R&D race, in case of patenting we have

$$X_2^{P/NIC} - X_2^{P/NINC} = \frac{v}{2cr}$$

and

$$X_2^{P/NINC} - X_2^{P/I} = \frac{v}{2cr}.$$

Clearly, $X_2^{P/NIC} > X_2^{P/NINC} > X_2^{P/I}$.

In case of no patenting, we have

$$X_2^{NP/NIC} - X_2^{NP/NINC} = \frac{3\left(3\lambda - 2\mu\right)\mu^2 v - 2cr^2\lambda\left(\lambda - \mu\right)}{6cr\lambda^2\mu}$$

and

$$X_2^{NP/NINC} - X_2^{NP/I} = \frac{\mu^2 v + 6cr^2\left(\lambda - \mu\right)}{2cr\lambda\mu}.$$

Given the assumptions that aggregate investment by the rival firms in the second R&D race is positive and that $\lambda \geq \mu$, we get $X_2^{NP/NIC} > X_2^{NP/NINC} > X_2^{NP/I}$.

(ii) In the first R&D race, in case of patenting we have

$$X_1^{P/I} - X_1^{P/NINC} = \frac{2\lambda v - 3cr^2}{2cr\lambda}$$

and

$$X_1^{P/NINC} - X_1^{P/NIC} = \frac{r}{6\lambda}.$$

Given the assumption that aggregate investment by the rival firms in the second R&D race is positive, we get $X_1^{P/I} > X_1^{P/NINC} > X_1^{P/NIC}$.

In case of no patenting, we have

$$X_1^{NP/I} - X_1^{NP/NINC} = \frac{\mu v - 3cr^2}{cr\lambda}$$

and

$$X_1^{NP/NINC} - X_1^{NP/NIC} = \frac{(\lambda - \mu)\,\mu v}{cr\lambda^2} + \frac{r}{3\lambda}.$$

Again, given the assumptions that aggregate investment by the rival firms in the second R&D race is positive and that $\lambda \geq \mu$, we get $X_1^{NP/I} > X_1^{NP/NINC} > X_1^{NP/NIC}$.

D Proof of Proposition 3

It is necessary to show that $V_w^P > V_w^{NP}$ implies $W^P > W^{NP}$ under all three regimes. From (12) and (13), it is clear that $W^{P/I} > W^{NP/I}$ and $W^{P/NIC} > W^{NP/NIC}$ if $P\left(X_1^P\right) > P\left(X_1^{NP}\right)$ and $P\left(X_2^P\right) > P\left(X_2^{NP}\right)$ under the respective regimes. We have

$$P\left(X_1^{P/I}\right) - P\left(X_1^{NP/I}\right) = \frac{2cr^2\left[4cr^2\lambda - (\lambda-\mu)^2 v\right]}{(3\lambda v - 2cr^2)\left[(\lambda-\mu)^2 v + 3\lambda^2 v - 6cr^2\lambda\right]}$$

and

$$P\left(X_2^{P/I}\right) - P\left(X_2^{NP/I}\right) = \frac{2cr^2\,(\lambda-\mu)}{\lambda\mu v}.$$

Both of these expressions are positive given that $V_w^{P/I} > V_w^{NP/I}$ (which implies $4cr^2\lambda - (\lambda-\mu)^2 v > 0$), $\lambda \geq \mu$, and $X_0^{P/I} > 0$.

For regime NIC, we have

$$P\left(X_1^{P/NIC}\right) - P\left(X_1^{NP/NIC}\right) = \frac{6cr^2\left[4cr^2\lambda - 9\,(\lambda-\mu)^2 v\right]}{(3\lambda v + 2cr^2)\left[9\,(\lambda-\mu)^2 v + 3\lambda^2 v - 2cr^2\lambda\right]}$$

and

$$P\left(X_2^{P/NIC}\right) - P\left(X_2^{NP/NIC}\right) = \frac{2cr^2\,(\lambda-\mu)}{3\lambda\mu v}.$$

Again, both of these expressions are positive given that $V_w^{P/NIC} > V_w^{NP/NIC}$ (which implies $4cr^2\lambda - 9\,(\lambda-\mu)^2 v > 0$), $\lambda \geq \mu$, and $X_0^{P/NIC} > 0$.

Finally, for regime $NINC$, $W^{P/NINC} > W^{NP/NINC}$ if $P\left(X_1^{P/NINC}\right) > P\left(X_1^{NP/NINC}\right)$, $P\left(X_2^{P/NINC}\right) > P\left(X_2^{NP/NINC}\right)$, and $\left(\frac{\lambda X_o^{P/NINC}}{\lambda X_2^{P/NINC}+r}\right) > \left(\frac{\mu X_o^{NP/NINC}}{\lambda x_w^{NP/NINC}+\mu X_o^{NP/NINC}+r}\right)$. We can show in a similar fashion that all three inequalities hold given that $V_w^{P/NINC} > V_w^{NP/NINC}$ (which implies $cr^2\lambda - \left(3\lambda^2 - 4\lambda\mu + \mu^2\right) v > 0$) and $\lambda \geq \mu$.

E Proof of Proposition 4

(i) For $\mu \leq \mu^I$, we need to compare $W^{NP/I}$, $W^{NP/NIC}$, and $W^{NP/NINC}$. We have

$$W^{NP/NIC} - W^{NP/I} = \frac{2s}{r}\left[P\left(X_1^{NP/NIC}\right)P\left(X_2^{NP/NIC}\right) - P\left(X_1^{NP/I}\right)P\left(X_2^{NP/I}\right)\right].$$

Substituting for the relevant $P(X_1)$ and $P(X_2)$ functions we get

$$\frac{8crs(\lambda-\mu)\left[\left(48\lambda^2(\lambda-\mu)+9\mu^2(2\lambda-\mu)\right)v - 8cr^2\lambda(13\lambda-10\mu)\right]}{3\mu\left[\left(9(\lambda-\mu)^2+3\lambda^2\right)v-2cr^2\lambda\right]\left[\left((\lambda-\mu)^2+3\lambda^2\right)v-6cr^2\lambda\right]}.$$

Both the numerator and the denominator of this expression are positive given $X_o^{NP/I} > 0$ and that $\lambda \geq \mu$.

From the expressions for $W^{NP/NINC}$ and $W^{NP/NIC}$ it is clear that when $s = 0$, $W^{NP/NINC} > W^{NP/NIC} = 0$. Consider the expression for $\frac{W^{NP/NINC}}{W^{NP/NIC}}$.

$$\frac{W^{NP/NINC}}{W^{NP/NIC}} = \frac{P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right)}{P\left(X_1^{NP/NIC}\right)P\left(X_2^{NP/NIC}\right)} + \frac{P\left(X_1^{NP/NINC}\right)\left(\frac{\mu X_o^{NP/NINC}}{\lambda x_w^{NP/NINC}+\mu X_o^{NP/NINC}+r}\right)v}{P\left(X_1^{NP/NIC}\right)P\left(X_2^{NP/NIC}\right)2s}.$$

Taking the derivative with respect to s, we have $\frac{\partial\left(W^{NP/NINC}/W^{NP/NIC}\right)}{\partial s} < 0$. As $s \to \infty$, $\frac{W^{NP/NINC}}{W^{NP/NIC}} \to \frac{P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right)}{P\left(X_1^{NP/NIC}\right)P\left(X_2^{NP/NIC}\right)}$. We need to show that this ratio is < 1.

The difference $P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right) - P\left(X_1^{NP/NIC}\right)P\left(X_2^{NP/NIC}\right)$ is equal to

$$\frac{cr^2\left(Av^2 - 4cr^2\lambda Bv + 12c^2r^4\lambda^2\right)}{3\mu(2\lambda-\mu)^2 v^2\left[-9(\lambda-\mu)^2 v - 3\lambda^2 v + 2cr^2\lambda\right]} \tag{A1}$$

where $A = 3\left(16\lambda^4 - 40\lambda^3\mu + 28\lambda^2\mu^2 - 6\lambda\mu^3 + 3\mu^4\right)$ and $B = \left(8\lambda^2 - 14\lambda\mu + 11\mu^2\right)$. The denominator of this expression is negative assuming $X_o^{NP/I} > 0$ and $\lambda \geq \mu$. In the numerator, consider the first two terms inside the parenthesis. Since we are interested in the region where $\mu^I \geq \mu$, this implies $(\lambda - \mu)^2 v > 4cr^2\lambda$. We can easily verify that B is positive for all $\frac{\mu}{\lambda} \in [0, 1]$. This means multiplying both sides of $(\lambda - \mu)^2 v > 4cr^2\lambda$ by Bv would not affect the inequality. Thus, if we can show that $Av^2 > (\lambda - \mu)^2 Bv^2$, we know that the numerator of (A1) is positive. Evaluating $A - (\lambda - \mu)^2 B$ at all $\frac{\mu}{\lambda} \in [0, 1]$ shows that it is always positive. Thus, the ratio $\frac{P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right)}{P\left(X_1^{NP/NIC}\right)P\left(X_2^{NP/NIC}\right)} < 1$.

(ii) For $\mu^{NINC} > \mu > \mu^I$, we first compare welfare under I with welfare under NIC to show that regime I always dominates. Since $\mu^{NINC} > \mu^{NIC}$, we first consider $\mu^{NIC} > \mu > \mu^I$. We need to compare $W^{P/I}$ and $W^{NP/NIC}$. The denominator of $W^{P/I} - W^{NP/NIC}$ is equal to $3\lambda\mu v\left(3\lambda v - 2cr^2\right)\left[9(\lambda - \mu)^2 v + 3\lambda^2 v - 2cr^2\lambda\right]$, is positive given $X_o^{P/I} > 0$. In the numerator we have

$$2crs\left[9\lambda Cv^2 - 6cr^2Dv + 8c^2r^4\lambda(4\lambda - 3\mu)\right] \quad \text{(A2)}$$

where $C = \left(8\lambda^3 - 26\lambda^2\mu + 36\lambda\mu^2 - 15\mu^3\right)$ and $D = \left(16\lambda^3 - 35\lambda^2\mu + 42\lambda\mu^2 - 18\mu^3\right)$. We would like to show that this is positive. We do this by first checking that $C > 0$ and $D > 0$ for all $\frac{\mu}{\lambda} \in [0, 1]$. Next, we note that $\mu > \mu^I$ implies $4cr^2\lambda > (\lambda - \mu)^2 v$ and $X_o^{P/I} > 0$ implies $\lambda v > 4cr^2$. Using these two inequalities, we can substitute $4cr^2$ for λv in the first term and $(\lambda - \mu)^2 v$ for $4cr^2\lambda$ in the last term inside the brackets. We get

$$2crs\left[36cr^2Cv - 6cr^2Dv + 2cr^2(4\lambda - 3\mu)(\lambda - \mu)^2 v\right]. \quad \text{(A3)}$$

(A3) is smaller than (A2) since all we have done is to make the positive terms in (A2) smaller. Thus, if we can show that (A3) is positive, we know that (A2) is positive. Factoring out cr^2v and evaluating (A3) for all $\frac{\mu}{\lambda} \in [0, 1]$ shows that it is positive.

For $\mu^{NINC} > \mu > \mu^{NIC}$, we need to compare $W^{P/I}$ and $W^{P/NIC}$. We have $W^{P/I} -$

$W^{P/NIC} = \frac{4crs\left(3\lambda v - 4cr^2\right)^2}{3\lambda v\left(9\lambda^2 v^2 - 4c^2 r^4\right)}$, which is positive assuming $X_o^{P/I} = \left(\lambda v - 4cr^2\right)/2cr\lambda > 0$.

Finally, we compare $W^{P/I}$ with $W^{NP/NINC}$. From the expressions for $W^{NP/NINC}$ and $W^{P/I}$ it is clear that when $s = 0$, $W^{NP/NINC} > W^{P/I} = 0$. Consider the expression for $\frac{W^{NP/NINC}}{W^{P/I}}$.

$$\frac{W^{NP/NINC}}{W^{P/I}} = \frac{2P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right)}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]} + \frac{P\left(X_1^{NP/NINC}\right)\left(\frac{\mu X_o^{NP/NINC}}{\lambda x_w^{NP/NINC}+\mu X_o^{NP/NINC}+r}\right)v}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]s}.$$

Taking the derivative with respect to s, we have $\frac{\partial\left(W^{NP/NINC}/W^{P/I}\right)}{\partial s} < 0$. As $s \to \infty$, $\frac{W^{NP/NINC}}{W^{P/I}} \to \frac{2P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right)}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]}$. We would like to show that this ratio is < 1.

We first note that $\frac{P\left(X_1^{P/NINC}\right)\left[1+P\left(X_2^{P/NINC}\right)\right]}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]} = 1 + \frac{35}{14}cr^2\left(\frac{\left(\lambda v-2cr^2\right)}{\left(\lambda v+cr^2\right)\left(-3\lambda v+4cr^2\right)}\right) < 1$ since the second term is negative for $X_o^{P/I} > 0$. Therefore, it is sufficient to show that $2P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right) < P\left(X_1^{P/NINC}\right)\left[1+P\left(X_2^{P/NINC}\right)\right]$ since we know that $P\left(X_1^{P/NINC}\right)\left[1+P\left(X_2^{P/NINC}\right)\right] < P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]$. Substituting for the relevant $P\left(X_1\right)$ and $P\left(X_2\right)$ values we get

$$\frac{cr^2\left[\lambda\left(8\lambda^3 - 24\lambda^2\mu + 22\lambda\mu^2 - 5\mu^3\right)v^2 + cr^2\left(4\lambda^3 - 2\lambda\mu^2 + \mu^3\right)v - 4c^2r^4\lambda^2\right]}{\lambda\mu\left(2\lambda-\mu\right)^2 v^2\left(\lambda v + cr^2\right)}.$$

The denominator is clearly positive. In the numerator, $2cr^2\lambda^3\left(\lambda v - 2cr^2\right) > 0$ for $X_o^{NP/NINC} > 0$. To see that the rest of the terms are positive, note that we are interested in $\mu > \mu^I$ which implies $4cr^2\lambda > \left(\lambda - \mu\right)^2 v$. Proceeding as we did while trying to show that $W^{P/I} > W^{NP/NIC}$, we can use this inequality to make the positive term smaller and to see that the expression we get is still positive for $\lambda \geq \mu$.

(iii) For $\mu \geq \mu^{NINC}$, we know that $W^{P/I} > W^{P/NIC}$ from part (ii). From the expressions for $W^{P/NINC}$ and $W^{P/I}$ it is clear that when $s = 0$, $W^{P/NINC} > W^{P/I} = 0$. Consider the expression for $\frac{W^{P/NINC}}{W^{P/I}}$.

$$\frac{W^{P/NINC}}{W^{P/I}} = \frac{P\left(X_1^{P/NINC}\right)\left[1+P\left(X_2^{P/NINC}\right)\right]}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]} + \frac{P\left(X_1^{P/NINC}\right)\left(\frac{\lambda X_o^{P/NINC}}{\lambda X_2^{P/NINC}+r}\right)v}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]s}.$$

Taking the derivative with respect to s, we have $\frac{\partial\left(W^{P/NINC}/W^{P/I}\right)}{\partial s} < 0$. As $s \to \infty$, $\frac{W^{P/NINC}}{W^{P/I}} \to \frac{P\left(X_1^{P/NINC}\right)\left[1+P\left(X_2^{P/NINC}\right)\right]}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right]} = 1 + \frac{35}{14}cr^2\left(\frac{\left(\lambda v - 2cr^2\right)}{\left(\lambda v+cr^2\right)\left(-3\lambda v+4cr^2\right)}\right)$. This ratio is less than 1 since the second term is negative for $X_o^{P/I} > 0$.

F Proof of Proposition 5

Note that s^{high} is independent of μ since it is the critical s value for $\mu \geq \mu^{NINC}$. We have:

$$s^{high} = \frac{P\left(X_1^{P/NINC}\right)\left(\frac{\lambda X_o^{P/NINC}}{\lambda X_2^{P/NINC}+r}\right)v}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right] - P\left(X_1^{P/NINC}\right)\left[1+P\left(X_2^{P/NINC}\right)\right]}$$

and

$$s^{int} = \frac{P\left(X_1^{NP/NINC}\right)\left(\frac{\mu X_o^{NP/NINC}}{\lambda x_w^{NP/NINC}+\mu X_o^{NP/NINC}+r}\right)v}{P\left(X_1^{P/I}\right)\left[1+P\left(X_2^{P/I}\right)\right] - 2P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right)}.$$

It suffices to show that the numerator of s^{high} is greater than the numerator of s^{int} and that $P\left(X_1^{P/NINC}\right)\left[1+P\left(X_2^{P/NINC}\right)\right] > 2P\left(X_1^{NP/NINC}\right)P\left(X_2^{NP/NINC}\right)$. We have already shown that the second inequality holds in the Proof of Proposition (4), part (ii). To see that the first inequality holds, note that its denominator is equal to $2\lambda\mu\left(2\lambda-\mu\right)^2 v\left(\lambda v+cr^2\right)$, which is always positive. The numerator is given by

$$\begin{aligned}&\lambda\mu\left(\lambda-\mu\right)\left(2\lambda-\mu\right)^2 v^3 + cr^2\left(8\lambda^4 - 16\lambda^3\mu + 8\lambda^2\mu^2 + 2\lambda\mu^3 - \mu^4\right)v^2\\ &+c^2r^4\left(4\lambda^3 - 4\lambda^2\mu + \mu^3\right)v - 4c^3r^6\lambda^2\end{aligned}$$

which is positive assuming $X_o^{NP/NINC} > 0$ and $\lambda \geq \mu$.