Optimal Standards of Proof for Non-Obviousness and Infringement

Ezra Friedman and Abraham L. Wickelgren

June 4, 2018

Abstract

We build a model of patent protection under two forms of uncertainty; uncertainty regarding whether the original invention merits protection (non-obviousness), and uncertainty as to whether a particular competitor's product should be barred (infringement). We find that when it is practical to increase the rewards from innovation by extending patent length, the standards of proof for non-obviousness should be high. The intuition for this is that it is when it is practical to extend patent length, patent length should be set so that the increase in innovation from extending patent length is balanced by the increase in deadweight loss by extended monopoly pricing. In this situation, the ex-ante cost of failing to protect a good patent is minimal, but there is substantial deadweight loss from protecting a bad patent. In contrast, if non-infringing competing inventions substantially decrease the original inventor's profits, it might be desirable to have a very low standard proof for infringement.

1 Introduction

The standard of proof in patent litigation is an area which has not received much attention from economic theorists. In this paper we develop a model of innovation and patent enforcement and introduce two forms of uncertainty. We first consider uncertainty regarding non-obviousness. Under patent doctrine, an innovation can be patented only if it is non-obvious. From an economic modeling point of view, we regard this as a requirement that the innovation is of a type that requires some intentional investment to develop, as opposed to an invention that can be developed at zero cost (or whose discovery is entirely independent of inventive effort). From an economic standpoint, the supply of obvious inventions does not depend on patent protection, so there is no justification for the deadweight loss that would result from patent protection.

The next type of uncertainty that we consider is uncertainty regarding infringement. To study the optimal standard of proof for infringement, one needs to develop an economic theory of why some products should be blocked by a patent, while others should be ruled non-infringing. We build on previous literature, such as Gilbert and Shapiro(1990), that has identified patent breadth as a policy instrument that can be used to moderate the tradeoff between innovation incentive and deadweight loss from restricting innovation. As explained by Denicolo (96), building on the results from Klemperer(1996), the desirability of broad or narrow patents depends on a comparison between the rate at which increasing breadth increases the innovator's profits, compared to the rate at which increasing breadth decreases social surplus. If increasing breadth decreases inefficient copying or inefficient substitution, broad patents tend to be optimal. On the other hand, if decreasing breadth only slightly constrains the innovator's ability to price at the monopoly level, and thus reduces deadweight loss much more than profits, narrow and long patents may be optimal. Following this intuition, we see the aim of an infringement ruling as determining whether entry by the product in question would increase total surplus enough to justify the reduction of the original inventor's profit.

We find that when patent length is not constrained by practical concerns about deterring follow on innovation, it is more efficient to increase the rewards to innovation by increasing patent length than by reducing the standard of proof of non-obviousness. In this situation, either the optimal patent length is infinite, or the standard of proof for non obviousness should be certainty. The intuition behind this result is that when patent length is unconstrained, it should be set so that the marginal benefit from innovation induced by extending the patent is equal to the deadweight loss from extending the patent. This implies that there is no ex-ante welfare loss from a marginal decrease in the likelihood the invention receives patent protection. On the other hand, granting patent protection to an obvious invention creates deadweight loss from monopoly pricing without increasing incentives to innovate. Therefore, if there is any doubt as to whether an invention is non-obvious, patent protection should not be granted. This result generalizes to a result that in the absence of concerns regarding follow up innovation, infinite patent lengths are optimal unless some inventors can prove with certainty that their inventions are non-obvious.

We do not interpret this result as implying that patent lengths should be infinite, or even longer, in the real world. We do think that it is a theoretical contribution to isolate the relative desirability of using different policy levers to adjust rewards to patents. At a minimum, if there is slack in the other policy levers, it is desirable to insist upon a high standard of proof of obviousness.

A patent scheme that gives large rewards to only the most conclusively non-obvious inventions, and no reward otherwise is analogous to the (nearly) optimal incentive scheme described by Mirrlees(1974). When rewarding (or punishing) is socially costly, it is optimal to reserve the incentives for the cases which are most informative about the behavior one is trying to reward or punish. Focusing on the legal context, Louis Kaplow [cite] uses this intuition to show that when punishment is costly, it may be preferable to use extreme punishments with a high standard of proof. Because awarding patent protection is socially costly, it is most efficient to reserve the awarding of a patent for the cases where it is most likely to create an incentive to innovate.

On the other hand, as shown by Holmstrom and Milgrom (1987), if an agent has private information, they may be able to game a non-linear award scheme, and provide the most effort when it is most likely to lead to an award, as opposed to when it is most likely to be socially useful. As a result, when the innovator has private information about how the patent is likely to be judged, it may be less desirable to demand a very high standard of proof. We do find that introducing private information moderates our results somewhat, and our result that it is always optimal to use infinite patent length no longer holds. We engage in some simulation analysis where inventors have private information, and our results suggest that it is most likely to be desirable to use shorter patents and a lower standard of proof if there are very few obvious inventions. Thus, the posterior likelihood that an obvious invention is granted patent protection would still be fairly low. Furthermore, our analysis suggests that when there is sufficient private information, a reward scheme that is non monotonic in the courts's perception of non-obviousness may be superior to a monotonic scheme with an interior standard of proof for non-obviousness. Although we acknowledge that such a non-monotonicity may not be practical, we interpret this result as further evidence for the robustness of the optimality of a corner solution.

In addition to the standard of proof regarding obviousness, we construct a model to examine the optimal standard of proof regarding infringement. In order to model doubt about infringement, we build on the results of Denicolo(1996) regarding optimal scope of patents . The general results of that literature is that the optimal breadth is determined by looking at the ratio between decreased deadweight loss from narrowing the patent to the decrease in inventor's profit from broadening the patent, and comparing that ratio with the ratio of the decrease in deadweight loss from shortening the patent to term to decrease in inventor's profit from a shorter term. We model an infringing product as one that would be within the optimal scope of the patent, so allowing it to be marketed would decrease inventor's profits substantially, for a relatively small increase in total ex-post surplus. Marketing a non-infringing entrant's product would lead to a large increase in total surplus, relative to the decrease in the original inventor's profits.

In contrast to the results regarding obviousness, we do not find that courts should always require a high standard of proof for infringement, even if patent length is unconstrained. Intuitively, the courts should rule a new product infringes on the patent if the increase in the original inventor's profits from finding infringement divided by the increase in deadweight loss is greater than the expected increase in a non-obvious inventor's profit divided by the increase in deadweight loss that would come from finding the marginal invention non-obvious. In contrast to granting a patent to an obvious invention, finding infringement always increases the incentive to innovate. Thus, there is some positive side effect from an erroneous finding of infringement. We note that in this simple two-type model, it is possible that the standard of proof drops below zero. The intuition here is that when patent length is constrained, increasing the rewards to invention by broadening the patent can be more efficient than increasing the rewards to invention by reducing the standard of proof for non-obviousness. Thus, doubt regarding the validity of patents could lead to an optimal breadth that is broader than it would be under certainty.

In the next section, we describe our model of innovation with an imperfect signal of non-obviousness. We then extend the model by adding private information regarding how a court is likely to perceive an innovation. The next section describes a model of uncertain infringement in the context of uncertainty regarding non-obviousness.

2 Model of Non-Obviousness

There is an inventor who chooses work effort x in period 0. With probabilities depending on work effort, the effort results in non-obvious invention, an obvious invention, or no-invention at all. The probability of an obvious invention is given by a constant q. Thus, unlike Yelderman[2017], we assume that making an obvious invention more remunerative does not have any impact on the incentive to create a non-obvious invention. The probability of a non-obvious invention is given by the thrice differentiable function $f(x)$. We assume $f\prime(x) > 0$. $f\prime\prime(x) < 0$, and $lim_{x\to\infty} f(x) < 1 - q$ (so that the probability of any invention is always less than one). We also assume that $\frac{f\prime(x)}{-f""""(x)}$ is decreasing, which helps us obtain uniqueness. Finally, we assume that the obvious and non obvious inventions are mutually exclusive, so the likelihood of no invention at all is simply $1 - f(x) - q$.

For any invention, the court observes a signal $s \in [0, 1]$ of the invention's non-obviousness. One can think of this signal as a summary measure of all the information the court receives about obviousness. Let us use Φ to represent the cumulative distribution function over s for a non-obvious invention. Let ϕ be the probability density of the signal. The cumulative distribution, and probability density for the signal from an obvious invention is given by Ψ and ψ respectively. We assume that $\frac{\phi(s)}{\psi(s)}$ satisfies the monotone likelihood ratio property (MLRP). Without further loss of generality, we assume that s is defined such that $\frac{\phi(s)}{\psi(s)} = \frac{s}{1-s}$ is the likelihood ratio. Thus a rational court, expecting effort x would assess the probability that an observed invention was non obvious as $\frac{sf(x)}{sf(x)+(1-s)q}$. Note that this implies some restrictions on Φ. Specifically $\int_0^1 \phi(s)ds < 1$ and $\int_0^1 \frac{(1-s)\phi(s)}{s} ds \leq 1$

We assume that ϕ is positive and bounded over $(0, 1)$, but allow that there may be an atom in ϕ at 1. This implies that there may be a discrete probability that the inventor is able to show with certainty that the invention is non-obviousness. We also allow that there might be strictly positive probability of a signal $s = 0$ if the invention is obvious.

We define a simple threshold patent policy as a pair $(\tilde{s}, T)$ where T is patent length, and $\tilde{s}$ is a standard of proof of non-obviousness. We assume that $0 < T \leq \hat{T}$, where $\hat{T} \in (0, \infty]$. We are thinking of $\hat{T}$ as a reduced form abstraction of the constraints on patent length. Effective patent length may be constrained by the availability of more advanced technology, or by changing tastes, or by practical concerns such as a reluctance to impede cumulative innovation.[1]

We assume that the welfare consequences of an invention depend on the patent protection, and for now, we assume they are the same regardless of whether it is obvious or not (in the simple model, q parameterizes the social welfare consequences of protecting obvious inventions). If an invention is patented, it leads to flow profits of π_M for the inventor, and flow consumer surplus of C_M. If the invention is not patented it leads to profits of 0, and flow consumer surplus of C_C. In section 3, we will consider intermediate levels of patent protection that may allow competing products that reduce, but do not eliminate the inventor's profits. We assume that $\pi_M < C_C - C_M$, so patent protection is *ex post* inefficient.

We assume that there is a constant common discount factor β. The value of a patent to the inventor is $\int_0^T \pi_M \beta^t dt = \frac{\pi_M(1-\beta^T)}{-\ln \beta}$. We define this as $\Pi(T)$. Likewise, total consumer surplus is $\int_0^T C_M \beta^t dt + \int_T^\infty C_C \beta^t dt = \frac{\beta^T C_C + (1-\beta^T) C_M}{-\ln \beta}$, defined here as $C(T)$. So $C(T)$ is the consumer surplus from a non obvious invention that is granted patent protection. The *ex post* cost of the patent to consumers, in terms of consumer welfare loss is $\frac{(C_C - C_M)(1-\beta^T)}{-\ln \beta}$, for the sake of brevity, we refer to this as $D(T)$. Note that the social value of a non obvious invention that is not granted patent protection is $\frac{C_C}{-\ln \beta}$ or $D(T) + C(T)$. Thus $D(T) - \Pi(T)$ is the ex post welfare loss from patenting an invention.

The expected payoff to the inventor is thus given by

$$u(x) = f(x)(1 - \Phi(\tilde{s}))\Pi(T) + q(1 - \Psi(\tilde{s}))\Pi(T) - x \quad (1)$$

An inventor chooses x to maximize this. Therefore, the choice of x satisfies the following first order condition

$$f\prime(x)(1 - \Phi(\tilde{s}))\Pi(T) = 1 \quad (2)$$

Now, let us consider the social welfare consequences of either extending the patent term (T) or lowering the standard of proof. First, the inventor's first order condition

[1] More detailed consideration of the interaction between patent length, standard of proof, and cumulative innovation is a subject for another paper

(2) implies that the consequences of extending the patent term on effort are given by

$$\frac{dx}{dT} = \frac{\beta^T \pi_M}{-f''(x)(1-\Phi(\tilde{s}))\Pi(T)^2} \tag{3}$$

Since the inventor is internalizing the effects of her effort on her profits, and she is maximizing her profits with respect to her effort, the welfare consequences of an increase in the inventor's effort is simply the change in expected consumer welfare from inventions ($C(T)+\Phi(\tilde{s})D(T)$). The ex post welfare effect holding effort constant is the dead weight loss from patent protection for one extra period, T periods in the future, ($\beta^T(C_C - C_M - \pi_M)$) times the probability of patent protection in the first place ($f(x)(1-\Phi(\tilde{s}) + q(1-\Psi(\tilde{s}))$). The net welfare effect of extending the patent term is:

$$\frac{dW}{dT} = \frac{dx}{dT} f'(x)(C(T) + \Phi(\tilde{s})D(T)) - \beta^T[f(x)(1-\Phi(\tilde{s}) + q(1-\Psi(\tilde{s}))](C_C - C_M - \pi_M) \tag{4}$$

Let us consider the effect of changing $\tilde{s}$, the standard of proof. First, the inventor's first order condition (2) implies that the effect on effort of increasing the standard of proof is given by

$$\frac{dx}{d\tilde{s}} = \frac{\phi(\tilde{s})}{f''(x)(1-\Phi(\tilde{s}))^2\Pi(T)} \tag{5}$$

Note that term is negative because $f''(x) < 0$. So the welfare consequences are given by:

$$\frac{dW}{d\tilde{s}} = \frac{dx}{d\tilde{s}} f'(x)(C(T) + \Phi(\tilde{s})D(T)) + \phi(\tilde{s})(f(x) + q\frac{1-\tilde{s}}{\tilde{s}})(D(T) - \Pi(T)) \tag{6}$$

The first term is the social loss from the reduction in the probability of invention due to decreased effort. The second term is the ex post social gain from more competition. The term $\phi(\tilde{s})(f(x) + q\frac{1-s}{\tilde{s}})$ represents the reduction in the probability that a patent is granted as a result in the increase in standard of proof; $D(T) - \Pi(T)$. is the associated ex post social gain. This allows to establish the following result:

Proposition 1 *With only one type of inventor, either the optimal patent length is as long as possible ($T = \hat{T}$), or the optimal standard of proof is certainty ($\tilde{s} = 1$)*

Proof. All proofs are in the appendix. ■

Intuitively, strengthening patent protection by either extending the term, or by lowering the standard of proof, increases the incentive to invest in innovation at the cost of more deadweight loss from monopolies. However, the benefits of extending the patent term are more concentrated in non-obvious inventions, so per unit of deadweight loss created, extending the length encourages more innovation. The reason for this is that increasing the patent length increases the reward to the average valid patent, while decreasing the standard of proof increases the reward to the marginally valid, marginally obvious patent. Since the average valid patent is more likely to be the result of genuine innovative activity than the marginal patent, increasing the length steers proportionally more rewards to genuine innovative activity. The only exception occurs when there is a conclusive signal that occurs frequently enough to provide sufficient incentives with finite patent length.

2.1 Private Information and Multiple types

As mentioned before, incentive schemes with large, infrequent rewards can perform poorly in the presence of private information. Suppose that inventors have private information regarding the likelihood that their non-obvious invention is seen as obvious. Specifically, we an additional type of inventor, type L. We assume that the inventor's type is private information for the inventor, and that signal distribution depends on the inventor's type. The ex ante probability of the original type of inventor is λ. When the original (which we will sometimes refer to as the high type) type develops a non-obvious invention, it is more likely that there is a strong signal of non-obviousness than when the low type develops a non-obvious invention. Thus, while a very high cutoff for non-obviousness coupled with a very long patent term might be optimal for only one type, one might conjecture that an interior solution is could be optimal with multiple types.

We analyze this formally by adjusting the model as follows. The probability distribution over the signal from a non-obvious invention is given by Φ_L, with the density given by ϕ_L, for low types. We assume a Monotone Likelihood Ratio Property(MLRP), so $\frac{\phi(s)}{\phi_L(s)}$ is increasing in s, and $\frac{1-\Phi(\bar{s})}{1-\Phi_L(\bar{s})} > \frac{\phi(s)}{\phi_L(s)} > \frac{\Phi(s)}{\Phi_L(s)}$for any $s < \bar{s}$. We assume that an obvious invention generates the same signal distribution, with density and cumulative probability given by $\psi(s)$ and $\Psi(s)$, respectively, regardless of which type invents it.

The analysis for two types follows the same steps as for one type. The analogous expression to (4) for the effect on total welfare from a change in patent length with two types is:

$$\lambda\frac{dx}{dT}f\prime(x)(C(T)+\Phi(\tilde{s})D(T))+(1-\lambda)\frac{dx_L}{dT}f\prime(x_L)(C(T)+\Phi_L(\tilde{s})D(T))$$
$$-\beta^T\{\lambda f(x)(1-\Phi(\tilde{s})+(1-\lambda)f(x_L)(1-\Phi_L(\tilde{s})+q(1-\Psi(\tilde{s})))\}(C_C-\pi_M-C_M) \quad (7)$$

The analogous expression to (6) for the effect on total welfare from a change in the standard of proof with two types is:

$$\frac{dW}{d\tilde{s}}=\lambda\frac{dx}{d\tilde{s}}f\prime(x)(C(T)+\Phi(\tilde{s})D(T))+(1-\lambda)\frac{dx_L}{d\tilde{s}}f\prime(x_L)(C(T)+\Phi_L(\tilde{s})D(T))$$
$$+\{\lambda f(x_H)\phi(\tilde{s}+(1-\lambda)f(x_L)\phi_L(\tilde{s})+q\psi(\tilde{s})\}[D(T)-\Pi(T)] \quad (8)$$

Our first result is that, holding patent length fixed, it would be desirable to use a lower threshold once we introduce the low type who has less accurate signals of non-obviousness. Because it is more difficult for the low type to prove the invention is non-obvious, low type inventors have insufficient incentive to invest unless the standard of proof is lowered.

Lemma 1 *Suppose that patent threshold is optimal for original types so that (6)=0 and* $\lambda < 1$*, then* $\frac{dW}{d\tilde{s}} < 0$*, so the threshold is above the optimal level for mixed types.*

Proof. See Appendix ■

This leaves the question of if and when finite patent lengths might be optimal. A simple corollary of proposition (1), is that as $\lambda \to 1$ or $\lambda \to 0$, the optimal patent length is infinite. Obviously, as $\lambda \to 1$, we are in the one-type situation, addressed above, but as $\lambda \to 0$ we are again in a one type situation, albeit with less informative signals. But since the proof of proposition (1) does not depend on the informativeness of the signal, proposition (1) still applies.

Our next result suggests that patent length below the maximum can only be optimal only if the optimal policy gives too much incentive to high-signal types without considering the possibility of accidentally rewarding obvious inventions. In other words, in order for patent length below the maximum to be optimal, the policy must give more expected reward to high type with non-obvious inventions than an optimal policy which could perfectly identify non-obvious inventions would.

In order to present this result formally, we reduce notation by using the following change of variables:

$$S_H=\frac{f\prime(x)(C(T)+\Phi(\tilde{s})D(T))}{-f\prime\prime(x)(1-\Phi(\tilde{s}))^2\Pi(T)^2}$$

and

$$S_L=\frac{f\prime(x_L)(C(T)+\Phi_L(\tilde{s})D(T))}{-f\prime\prime(x_L)(1-\Phi_L(\tilde{s}))^2\Pi(T)^2}$$

These roughly represent the ex-ante marginal social value of increasing the expected profit from a non obvious inventor of each type.

Lemma 2 *Suppose that the optimal policy is the pair* $(\tilde{s}, T)$ *and* $T < \hat{T}$*, so that a less than maximal length patent is optimal. Let* $W_C = C_C$ *and* $W_M = C_M + \pi_M$ *be the per period social welfare gains from the innovation. Then* $S_H \pi_M < f(x)(W_C - W_M)$*, implying that the value of additional innovation gained by increasing the rewards to high-signal type non-obvious inventors is outweighed by the deadweight loss from providing those rewards.*

Proof. See Appendix. ■

Let $s^\dagger$ be the optimal non-obviousness signal threshold when $T = \hat{T}$, holding all else constant. That is to say $s^\dagger$ is the optimal threshold signal when patent length is maximized. Let $x^\dagger$ be the level of effort for a high type inventor when $T = \hat{T}$, and the threshold signal is $s^\dagger$. We derive the following sufficient condition for maximal length to be optimal. The condition essentially looks at a scenario of maximal patent length, and the optimal signal threshold conditional on maximal length, and asks whether it would be beneficial to increase the likelihood that the high type non-obvious inventors obtain patents without affecting the likelihood of patents for obvious inventions or low type inventors.

Proposition 2 *Suppose* $\Pi(\hat{T})\frac{-f\prime(x^\dagger)}{f\prime\prime(x^\dagger)}(V(\hat{T})+\Phi_H(s^\dagger)C(\hat{T}))-f(x^\dagger)(D(\hat{T})-\Pi(\hat{T})) > 0$*, then* $T^* = \hat{T}$

If we'd like to increase the patent rewards to high type inventors under the optimal policy, then patent length should be extended to the extent possible. We only have an interior solution if the high types have too much incentive to invent, even if we aren't worried about the deadweight loss from obvious patents. The intuition here is that lowering the standard of proof reduces the ratio of expected rewards from innovation between the accurate signal high type and less accurate signal (low) type. Thus, this can be ameliorated by lowering the standard of proof, and decreasing the patent length. This comes at the expense of increasing the deadweight loss from obvious inventions, so it is only likely to obtain when the are relatively few obvious inventions.

In order to assess under what circumstances patent length less than maximal would be optimal, we conducted simulations, and numerically solved for optimum patent length and signal threshold. We varied the elasticity of effort, the relative signal strength of the low type, and the relative proportion of low types. We calibrated the profit of a patented invention so that a patent life of about 18 years would be optimal in the absence of any obvious inventions.[2] We found that when there were two types of non-obvious inventors, there was generally a threshold level of obvious inventions q, below which non- maximal patent length is optimal.

We found that the maximum likelihood of obvious patents at which non-maximal patent length becomes optimal is generally quite low. Because this implies that inventions are ex-ante unlikely to be obvious, this implies that the posterior probability that a patent would be obvious at the optimal threshold is still low. Figure 1 shows the standard of proof (z) and the posterior likelihood that a patented invention is non-obvious (θ), assuming the maximal number of obvious inventions for which finite patent length is optimal. The x-axis is ζ, the ratio of the likelihood of the strongest signal of non-obviousness between the low type and the high type. For example, even when a non-obvious invention by the low type is half as likely to generate the strongest signal of non-obvious invention, compared to the high type, we found that the posterior likelihood that a patent was non-obvious at the signal threshold was above 90% .[3] Even when the low type was only one quarter as likely to generate the high signal, the posterior likelihood at the threshold was above 75%. See Figure 1. We note that as q, the likelihood of obvious invention decreases further, the optimal patent length gets shorter, and the signal threshold for non-obviousness gets lower, but the posterior probability that a patent is obvious at the signal threshold gets smaller as well. Although this admits the possibility of finite patent length, this result reinforces the intuition that the threshold for non-obviousness should be set so that few obvious inventions receive patents. We have finite patent lengths only when the posterior on non-obviousness is high, albeit because the prior was high to begin with. On the other hand if we have a lot of obvious inventions, we are more concerned with the tradeoffs between obvious and non-obvious inventions and are less concerned that there is too much incentive to invent any type of invention.

2.2 Additional policy levers

Some commentators have suggested that in addition to imposing a non-obviousness threshold it might be beneficial to vary the patent length according to characteristics of the patent. In particular, one might desire a policy that adjusts according to the estimated 'non-obviousness' of the patent, with shorter patent length for marginally obvious inventions. Our next result suggests that this intuition may be reversed, if

[2] To be precise, we assumed that $f(x)$, the probability of a non-obvious invention was given by $f(x) = x^{\alpha}$, with $\alpha < 1$. Thus a higher α implies more elasticity of effort. We assumed that $\phi(s) = s$ and $\psi(s) = 1 - s$. The signal density for the low type was $\psi_L(s) = \zeta s + (1-\zeta)(1-s)$, implying that the relative likelihood of the most informative signal between the high type and low type $\frac{\phi(1)}{\phi_L(1)} = \zeta$. We arbitrarily assumed that $C_C = 40$ and $C_M = 10$ and we set $\pi_M = (1 - \frac{5\alpha}{3})C_C - C_M$ so that optimal patent length would be about 18 years with a discount rate of a 5%.

[3] The assumptions for this simulation were that $\alpha = .25$ so $f(x) = x^{(1/4)}$, $\lambda = .5$, and $\pi_M = 8.3333$.

non-maximal patent length is optimal under a uniform threshold for obviousness, it can be more efficient to reward marginally non-obvious inventions while granting no protection to inventions which have a very strong signal of non-obviousness. Specifically, in the two-type model, rather than a constant, non-maximal, reward above a certain threshold, there is some pair $(\tilde{s}, \hat{s})$ such that it would be more efficient to offer infinite length patents to all inventions with a signal above $\tilde{s}$ and below $\hat{s}$.

Proposition 3 *Suppose that in the two type model, the optimal two lever policy is $(\tilde{s}, T)$, where $T < \hat{T}$ and $\tilde{s} < 1$. Then the two lever policy leads to lower social welfare than a policy triple, $(\tilde{s}, \hat{s}, \hat{T})$, which awards patents of maximal length $\hat{T}$ to inventions which produce a signal in the interval $(\tilde{s}, \hat{s})$, where $\tilde{s} < \hat{s} < 1$.*

This result comes about because finite length is only optimal if the high type is getting too much reward. By removing patent protection from the highest signals, and lengthening patent protection for the intermediate signals, we can reduce reward for the high type while keeping rewards for the low type constant. In practice, we think there would be practical difficulties with implementing such a policy. For example, it might be relatively easy for an applicant to weaken any signal of non-obviousness. Nonetheless, we do believe that this result demonstrates the difficulties with complex patent reward schemes; namely if the inventor possesses private information, it may be very difficult to tailor complex schemes towards inventions where incentives are likely to be necessary or valuable.

3 Infringement

In addition to patent length, previous literature has identified patent breadth as a policy instrument that can be used to moderate the tradeoff between innovation incentive and deadweight loss from restricting innovation. As explained by Denicolo (1996), building on the results from Klemperer(1990) and Gilbert and Shapiro(1990), the desirability of broad or narrow patents depends on a comparison between the rate at which increasing breadth increases innovators profits, compared to the rate at which increasing breadth decreases social surplus. If increasing breadth decreases inefficient copying, or inefficient substitution, broad patents tend to be optimal. On the other hand, if decreasing breadth allows competition at the margins which constrains the monopoly price and decreases deadweight loss proportionally more than it decreases monopoly profits, narrow and long patents are optimal.

This section considers the tradeoff between length, breadth, and certainty. We take the difference between infringing and non infringing products to be whether or not they would be within the optimal scope of the patent with optimal patent length under certainty. In other words, if the trade-off between the increased incentives to invent from excluding the product and the deadweight loss maintained by excluding the product is better than the trade off from extending the patent, the product is infringing. Thus, under conditions of certainty, a shorter patent that excluded the infringing product would be preferable to a longer patent that allowed the infringing invention. On the other hand, if the trade-off between the increased incentives to invent from excluding the product and the deadweight loss maintained by excluding the product is worse than the trade off from extending the patent, we'd prefer to allow the new product, and it is non-infringing.

3.1 Infringement model

If there is a patented invention, there is an ϵ chance[4] that there is an immediate opportunity for competitors to reverse engineer the invention and market a new product. For the purposes of this model, we assume that ϵ is small enough that we can disregard second order effects of the infringement enforcement policy, and thus we can evaluate the first order effects assuming that patent length and standard of proof for obviousness are unchanged from section 2.

With probability μ conditional on existing, the new product is infringing, and would lead to a relatively small increase in ex post total welfare if marketed. With an infringing product, flow profits to the original inventor are π_I, flow profits to the entrant copier are π_e, and consumer surplus is C_I. We assume $\pi_I < \pi_M$ (so entry harms the original inventor), and finally, we assume that $\frac{C_C-(\pi_M+C_M)}{\pi_M} > \frac{C_I+\pi_I+\pi_e-(\pi_M+C_M)}{\pi_M-\pi_I}$ so that excluding the infringing product is a more efficient way to increase an inventor's profits than extending the patent term. Following Denicolo, this implies that in conditions of certainty, it would not be socially desirable to broaden the patent to cover the infringing product, *unless* patent length was already infinite.

With probability $(1-\mu)$ the new product, if marketed, would increase social welfare substantially relative to the decrease in the inventors' profit. We call this product non-infringing, and assume that with a non-infringing product, flow profits to the original inventor are π_N, flow profits to the entrant copier are again π_e, and consumer surplus is $C_N > C_I$. In order to reduce complexity we assume that $\pi_N = \pi_I$, so the only difference between the infringing and non-infringing product is the impact on consumer welfare. Qualitatively, the effects of increasing π_N above π_I would be similar to the effects of a further increase in C_N, they would make it less attractive to erroneously find infringement from a non-infringing product. We use the symbol W_x to refer to flow welfare under the various competitive scenarios, so $W_M = C_M + \pi_M$, $W_N = \pi_N + \pi_e + C_N$, $W_I = \pi_I + \pi_e + C_I$, and also $W_C = C_C$. Note that taken together, these assumptions imply

$$\frac{\pi_N}{W_C - W_N} > \frac{\pi_M}{W_C - W_M} > \frac{\pi_I}{W_C - W_I}$$

[4] Increasing the likelihood of the new product does not change our intuition, but it makes the math less neat.

In terms of deadweight loss, allowing the non-infringing product is the most efficient way to reward the inventor, followed by monopoly, followed by allowing the infringing product.

3.2 Optimal Standard of Proof for infringement

To evaluate the optimal standard of proof for infringement, let us assume that there is a signal b of the likelihood that the new product is infringing. The cumulative distribution, and signal density of b is given by G_I and g_I respectively for infringing products. For a non-infringing product, the corresponding distribution functions are given by G_N and g_N respectively, and we assume a MLRP property so $\frac{g_I(b)}{g_N(b)}$ is increasing in b. Initially, we assume that the follow on product is costless to develop, so we can ignore the entrant's incentives to develop the new product, and there is no social loss from not being able to determine if a new product is infringing until it is produced. We briefly discuss the implications of considering the entrant's incentives in the discussion section.

Beginning with the optimal threshold for non-obviousness of the original patent, we can rewrite our first order condition on $\tilde{s}$ (6) as follows:

$$S\phi(\tilde{s})\Pi(T) = (\phi(\tilde{s})f(x) + q\phi(\tilde{s})\frac{1-s}{s})(D(T) - \Pi(T))$$

Define z as the posterior likelihood that an invention with $\tilde{s}$ is non obvious. Note $z = \frac{\phi(\tilde{s})f(x)}{\phi(\tilde{s})f(x)+q\phi(\tilde{s})\frac{1-s}{s}}$. Noting that $\frac{\Pi_{(T)}}{D(T)-\Pi(T)} = \frac{\pi_M}{W_C - W_M}$, we can rewrite this as:

$$\frac{z}{f(x)} = \frac{W_C - W_M}{S\pi_M}$$

Since we are not varying $\tilde{s}$, we will drop the arguments from the distribution functions for signals of obviousness, and for the remainder of this section we will use Φ, Ψ, ϕ , and ψ to refer to $\Phi(\tilde{s}), \Psi(\tilde{s}), \phi(\tilde{s})$ and $\psi(\tilde{s})$.

Since an infringing [non infringing] product has a $G_I(\tilde{b})$ $[G_N(\tilde{b})]$ of being ruled non-infringing, the expected return to the original inventor from a non obvious invention is

$$\frac{1-\beta^T}{-\ln\beta}(1-\Phi)\{\pi_M - \epsilon(\mu G_I(\tilde{b})(\pi_M - \pi_I) + (1-\mu)G_N(\tilde{b})(\pi_M - \pi_N))\}$$

The change in this term from increasing $\tilde{b}$ is given by $\frac{1-\beta^T}{-\ln\beta}(1-\Phi)\epsilon(\mu g_I(\tilde{b})(\pi_M - \pi_I) + (1-\mu)g_N(\tilde{b})(\pi_M - \pi_N))$, so the change in effort from increasing b is given by

$$\frac{dx}{d\tilde{b}} = \frac{1-\beta^T}{\ln\beta}(1-\Phi)\frac{\epsilon(\mu g_I(\tilde{b})(\pi_M - \pi_I) + (1-\mu)g_N(\tilde{b})(\pi_M - \pi_N))}{-f''(x)(1-\Phi(\tilde{s}))^2\Pi(T)^2}$$

The change in net value of innovation is thus

$$f'(x)\frac{dx}{d\tilde{b}}(C(T)+\Phi D(T) = \tag{9}$$

$$\frac{1-\beta^T}{-\ln\beta}S(1-\Phi)\epsilon(\mu g_I(\tilde{b})(\pi_M-\pi_I)+(1-\mu)g_N(\tilde{b})(\pi_M-\pi_N)) \tag{10}$$

The ex-post expected change in deadweight loss from restricted competition is:

$$(f(x)(1-\Phi)+q(1-\Psi))\epsilon\frac{1-\beta^T}{-\ln\beta}(\mu g_I(\tilde{b})(W_M-W_I)+(1-\mu)g_N(\tilde{b})(W_M-W_N))$$

Let y be the posterior that the follow on invention is infringing at $\tilde{b}$, so $y = \frac{\mu g_I(\tilde{b})}{\mu g_I(\tilde{b})+(1-\mu)g_N(\tilde{b})}$. Let $\theta = \frac{f(x)(1-\Phi(\tilde{s}))}{f(x)(1-\Phi(\tilde{s}))+q(1-\Psi(\tilde{s}))}$ be the posterior that a patented invention is non obvious. Note that $\theta > z$. Dividing both terms by $(f(x)(1-\Phi)+q(1-\Psi))\epsilon\frac{1-\beta^T}{-\ln\beta}(\mu g_I(\tilde{b})+(1-\mu)g_N(\tilde{b}))$, and substituting in our expression for the shadow value of innovation (S), we have our first order condition for the optimal infringement threshold:

$$S\frac{\theta}{f(x)}(y(\pi_M-\pi_I)+(1-y)(\pi_M-\pi_N)) = y(W_I-W_M)+(1-y)(W_N-W_M) \tag{11}$$

Before presenting our result regarding the optimal signal threshold, we introduce the symbol ρ to represent the relative effectiveness between complete protection, and narrow protection. Formally we define: $\rho = \frac{\pi_M-\pi_N}{W_N-W_M}/\frac{\pi_M}{W_C-W_M}$. Thus ρ is the ratio between the 'price' in terms of inventor's profits of reducing deadweight loss by narrowing the patent, and the 'price' of reducing deadweight loss by reducing patent term.

Proposition 4 *The optimal threshold signal $\tilde{b}$ of infringement should be set so that the posterior likelihood that a new product with signal $\tilde{b}$ is a me too invention, and thus infringing is given by*

$$y = \frac{(W_N-W_M)-S\theta(\pi_M-\pi_I)}{W_N-W_I} = \frac{W_N-W_M}{W_N-W_I}(1-\frac{\theta\rho}{z})$$

The term $\frac{\theta}{z}$ is a measure of the ratio between the probability the non-obviousness of the average patent granted to the probability of non-obviousness of the marginal patent granted. If the constraint on patent length is binding, then $\frac{\theta}{z} > 1$. On the other hand if the constraint on patent length is not binding, which could only happen if patents are awarded only to inventions with the highest signal of non-obviousness, then $\theta = z$. If $\frac{\theta}{z}$ is relatively large, implying that the marginal patent is relatively likely to be obvious, then increasing the reward for innovation from decreasing the standard of proof of non-obviousness is relatively inefficient. In this case it is more desirable to increase the rewards from innovation by decreasing the standard of proof for infringement. In fact, if ρ is high, so the ex-ante welfare gains from allowing the non-infringing product are low, we can have $y \leq 0$, implying that the entrant product should be barred, even when we were certain the product is 'non-infringing'. The intuition for this is that in a second best world, broadening patents beyond the optimal breadth can be a more efficient way to increase the rewards for invention than decreasing the standard of proof for obviousness, which would disproportionately increase the deadweight loss from patenting obvious inventions.

In contrast if ρ is very small, there is little negative effect in failing to enforce the patent, because allowing entry does not lead to a significant decrease in innovator profits relative to the potential decrease in deadweight loss. In that case, a higher standard of proof for infringement is likely to be optimal. The difference $W_M - W_I$ could be positive or negative, as W_I or W_N increase relative to W_M,, it becomes less attractive to enforce the patent against an entrant product so the standard of proof increases. If $W_I > W_M$, this implies that making a type 2 error on infringement is less harmful, because failing to find infringement has a positive side effect of increasing total welfare. Finally, as W_N draws close to W_I, there is less difference between the infringing and non-infringing product, so the optimal standard of proof becomes more sensitive to the other factors.

Proposition 5 *The optimal threshold posterior probability of infringement is: (i) decreasing in π_M; (ii) increasing in π_N, W_N and W_I (iii) increasing in W_M if and only if $z/\theta < (\pi_M - \pi_N)/\pi_M$. That is, a court should require a higher standard of proof for infringement if the monopoly profit of the innovator is lower when entry is blocked, if the innovator's profit given entry is higher and if total welfare from allowing entry of the non-infringing product or the infringing product is higher. The court should require a higher standard of proof when total welfare from blocking entry is higher if and only if the ratio of the posterior probability of non-obvious of the marginal patent to the average patent is sufficiently small.*

Proof. See Appendix ■

The larger the innovator's profits are without entry, the more costly it is, in terms of reducing innovation incentives, to allow the entrant's product in the market, so the lower standard of proof we want for an infringement ruling that keeps the entrant out of the market. By contrast, the larger the innovator's profits are with entry, the less entry reduces the innovation incentives, so the higher standard of proof we want for infringment. If the total welfare from letting in either an infringing or a non-infringing product is greater, the greater is the ex post cost of an infringement ruling that keeps out the entrant, which warrants a higher standard of proof. The effect of increasing the welfare when the entrant is excluded is more complicated. The direct effect is to reduce the cost of a wrongful finding of infringment, which suggests lowering the standard of proof. But, increasing the welfare without competition also tends to reduce the optimal standard of proof for non-obviousness, which decreases the need to increase innovation incentives through broadening the patent, suggesting a higher standard of proof for infringment.

4 Discussion

In contrast to questions of non-obviousness, we find that it is not generally optimal to demand a high standard of proof for an infringement claim. This result stems from the differential consequences of an erroneous finding in favor of the original inventor. Because an erroneous finding of infringement nonetheless encourages innovative activity, the imposition of additional deadweight loss has at least some ex-ante benefit. Of course if we assumed that $\pi_M = \pi_N$, this would no longer be the case, and the only difference between erroneous findings in favor of the original inventor would be of magnitude. However, we have good reason to believe in the generality of the assumption that $\pi_M > \pi_N$. To wit, the original inventor is unlikely to sue for infringement if a finding in their favor cannot increase their profits.

4.1 Incentives to enter

Our analysis of the optimal standard of proof for infringement has abstracted away from some important considerations, notably the incentives of potential competitors to develop new products. In general, whether or not an entrant has excessive or insufficient incentive to enter will depend on a comparison of the ex-post change in consumer welfare with the expected change in the original inventor's profit's, weighted by the shadow value of incentives for the original inventor. There will always be too much incentive to develop infringing products (since they are undesirable), but there may be excessive or insufficient incentive to develop non-infringing products. In particular, abstracting away from any litigation costs, an innovator who knows she has a non-infringing product will have too much incentive to develop the product if the private benefit from entry π_e is greater than the ex-ante social benefit from entry. The ex-ante social benefit is the difference between the decrease in deadweight loss $(W_N - W_M)$ and the expected net value of the decrease in innovation by the original inventor $(\theta(\pi_M - \pi_N))$. Thus there will be too much incentive to develop a non-infringing product if $\pi_e > W_N - W_M - S\theta(\pi_M - \pi_N)$, so if the entrant's profits are relatively high, the entrant has too much incentive to develop even non-infringing products. We note that our previous results imply that if y, the optimal standard of proof for infringement is positive, then $W_N - W_M > S\theta(\pi_M - \pi_N)$. Thus, as long as y is positive, as the entrant's profits approach zero, we know the entrant will never have socially efficient incentive to develop non-infringing products.

We can get some insight into how considering the investment incentives of the entrant will effect the optimal standard of proof for infringement through a simple extension to the model. Imagine that a potential entrant can choose between developing an infringing product or a non-infringing product, but it cannot develop both. Let the random variable $k \in [0, K]$, with distribution and density functions H and h, be the extra cost of developing a non-infringing product. The entrant will develop the non-infringing product if and only if $k < G_N(\bar{b})\pi_{eN} - G_I(\bar{b})\pi_{eI} \equiv \bar{k}$ where $\pi_{eN}(\pi_{eI})$ are the entrant's expected profits from entering with a non-infringing (infringing) product. Thus, the probability that the entrant will choose the non–infringing product is $H(\bar{k})$.

The social welfare gain from developing an non-infringing rather than an infringing product is

$$G_N(\bar{b})[W_N - W_M - S\theta(\pi_M - \pi_N)] - G_I(\bar{b})[W_I - W_M - S\theta(\pi_M - \pi_I)] - k =$$
$$\{G_N(\bar{b}) - G_I(\bar{b})\}[W_N - W_M - S\theta(\pi_M - \pi_N)] + G_I(\bar{b})[W_N - W_I + S\theta(\pi_N - \pi_I)] - k \tag{12}$$

The difference between the social welfare gain and the profit gain from developing the non-infringing rather than the infringing product is:

$$\{G_N(\bar{b}) - G_I(\bar{b})\}[W_N - \pi_{eN} - W_M - S\theta(\pi_M - \pi_N)] + G_I(\bar{b})[W_N - W_I - (\pi_{eN} - \pi_{eI}) + S\theta(\pi_N - \pi_I)] =$$
$$\{G_N(\bar{b}) - G_I(\bar{b})\}[C_N - C_M - (S\theta + 1)(\pi_M - \pi_N)] + G_I(\bar{b})[C_N - C_I + (S\theta + 1)(\pi_N - \pi_I)] \tag{13}$$

The second term is strictly positive. The first term is positive if and only if the social value of the non-infringing product exceeds the private value. Thus, there is insufficient incentive for the entrant to choose a non-infringing product whenever the social value of the non-infringing product exceeds the private value. There can also be insufficient incentive even if this is not the case if the probability of approval for an infringing product is not too small.

If there is insufficient incentives for the entrant to choose the non-infringing product, then considering the entrant's development choice will induce the social planner to adjust $\bar{b}$ to increase $\bar{k}$, the difference in the entrant's expected profit from developing a non-infringing versus an infringing product. To determine how this affects the optimal standard for infringement, we differentiate $\bar{k}$ with respect to $\bar{b}$ to get:

$$\frac{d\bar{k}}{d\bar{b}} = g_N(\bar{b})(\pi_{eN} - \pi_{eI}\frac{g_I(\bar{b})}{g_N(\bar{b})}) \tag{14}$$

By MLRP, we know that $\frac{g_I(\bar{b})}{g_N(\bar{b})}$ is increasing in $\bar{b}$, which means that there is a b^* such that if $\bar{b} < b^*$, then considering the entrant's project choice would push the optimal $\bar{b}$

up, while if $\bar{b} > b^*$, then considering the entrant's project choice would push the optimal $\bar{b}$ down. In other words, if the optimal standard of proof for infringement without considering the entrant's project choice decision were relatively low, considering the entrant's decision would suggest a higher standard in which it is harder to prove infringement. On the other hand, if the optimal standard of proof had been quite high, considering the entrant's project choice would suggest lowering that standard to make it easier to prove infringement.

4.2 Licensing

Our analysis to this point has assumed that if a new product is found infringing, the product is simply not produced, and the market outcome as if the the product never existed. However, parties often come to a licensing agreement as a result of a finding of infringement. We show that as long as licensing does not result in lower ex-ante welfare than no entry, the presence of licensing makes the consequences of a wrongful finding of infringement less detrimental, as well as making the consequences of a wrongful finding of non infringement worse, and thus the presence of licensing should lower the standard of proof for infringement. Normally the entry of new products is expected to increase social welfare, even if a monopolist controls the pricing of the new product along with the old, but exceptions can exist. Most notably, the new product might be a an inferior product that allows the monopolist to more effectively price discriminate if licensed. Since the welfare effects of price discrimination are ambiguous, it is possible that this price discrimination decreases social welfare.

Formally, let us assume that the prior symbols all referred to the market outcome under a finding of infringement without licensing and let us add L to the subscripts to refer to the market outcome when infringement is found, and the parties subsequently negotiate a licensing agreement.

First, we infer that $\pi_{NL} \geq \pi_M$ and $\pi_{IL} \geq \pi_M$, otherwise the original inventor would not agree to a license. We note that the new standard of proof should solve the first order condition.

$$S\theta(y_L(\pi_{IL} - \pi_I) + (1 - y_L)(\pi_{NL} - \pi_N)) = y_L(W_I - W_{IL}) + (1 - y_L)(W_N - W_{NL}) \quad (15)$$

The first side of the equality represents the benefit, in terms of incentive to innovate, from a finding of infringement. It is positive since $\pi_{IL} > \pi_M > \pi_I$ and $\pi_{NL} > \pi_M > \pi_N$. The second side is the cost, in terms of ex-post social welfare, of a finding of infringement.

Solving, we have:

$$y_L = \frac{W_{NL} - W_N + S\theta(\pi_{NL} - \pi_N)}{W_N - W_{NL} - W_I + W_{IL} - S\theta(\pi_{NL} - \pi_{IL})}$$

Proposition 6 *Suppose that licensing does not decrease social welfare, so $W_{NL} \geq W_M$ and $W_{IL} \geq W_M$, then $y_L \leq y$.*

Proof. Compare the first order condition with licensing (15) with the original, no-licensing, condition (11). Note that if $y_L = y$, then the left side, representing the benefit of finding infringement is greater with licensing. Furthermore, if $W_{NL} \geq W_M$ and $W_{IL} \geq W_M$, then the right side of (15) is less than or equal to the right side of (11), indicating that the cost, of an infringement finding in terms of lost social welfare, is no greater with licensing. Thus if y is the optimal standard of proof with no licensing, $S\theta(y(\pi_{IL}-\pi_I)+(1-y)(\pi_{NL}-\pi_N)) > y(W_I-W_{IL})+(1-y)(W_N-W_{NL})$, so the optimal standard of proof should be higher with licensing.

If we consider the entrant's project choice decision under licensing, then this effect is magnified. The reason for this is that under licensing, one would expect that the ratio of the entrant's profit increase from a finding of non-infringement for a non-infringing product relative to an infringing product would smaller than if licensing is not possible because there is probably a much smaller gain in combined profits from the infringing product than from the non-infringing product. This suggests that while a non-infringing product might receive a license that still leaves it with a reasonable share of its total profit, the infringing product might either receive no license at all or a license that leaves it with very little profit. If this ratio of the profit gain from a finding of non-infringement falls, then b^* falls, expanding the region under which considering the entrant's project choice incentives optimally leads to easier findings of infringement. ■

5 Conclusion

Our analysis shows that the optimal standard of proof for the non-obviousness of a patent is very different from the optimal standard of proof for infringement. In our base case, we find that the optimal standard of proof for non-obviousness should be very high combined with a very long patent term. This combination provides any given level of reward for valid inventions at the lowest expected patent length (and, thus, lowest dead weight loss) for obvious inventions (which will be invented without any patent protection). While this simple result is not fully robust to introducing private information among inventors about the quality of their signal of non-obviousness, we show that the intuitive force for this extreme scheme is still quite strong. That is, while interior solutions (non-maximal patent length) can be optimal with multiple types of inventors, this is only the case if such a scheme provides excessive innovation incentives for at least one type of inventor (ignoring the social cost of patenting obvious inventions). Furthermore, simulations show that even when an interior solution is optimal, the posterior probability of an obvious invention remains very small at the optimum.

Our analysis of infringement yields very different conclusions. For infringement, the central concern is that sometimes it is more efficient to reward innovation through a broad patent when a competing product would reduce profits a great deal relative to how much it would reduce dead weight loss, whereas for other products it is more efficient to have narrower but longer patent because the competing product would not reduce the innovator's profits much relative to its reduction in dead weight loss. Our problem for finding the optimal standard of proof is to balance these two considerations when we have only an imperfect signal of the nature of the competing product. There is no particular reason to think that the optimal standard of proof is extreme in this instance. We show that the optimal standard of proof is generally higher the greater the total welfare provided by a competing infringing or non-infringing product and the less either type of new product reduces the profits of the innovator. We also show that when we consider the possibility of licensing that the optimal standard of proof for infringement drops because the social consequences of wrongfully excluding a non-infringing product drops. We also show that when the entrant can choose between an infringing or a non-infringing product that this tends to make the optimal standard of proof less extreme because the incentive to choose the non-infringing product is maximized at an interior standard of proof for infringement. We leave further consideration of issues of cumulative innovation to future work.

6 Appendix

Proof. Proof of Proposition 1.

We will show that for any $\tilde{s} < 1$ and for any $T < \hat{T}$, either $\frac{dW}{d\tilde{s}} > 0$, or $\frac{dW}{dT} > 0$

Suppose that $\tilde{s} < 1$ and $\frac{dW}{d\tilde{s}} = 0$, we will show that $\frac{dW}{dT} > 0$. It will prove useful to compare $\frac{\beta^T \pi_M (1-\Phi(\tilde{s}))}{\phi(\tilde{s})\Pi(T)}$ with $\frac{\beta^T (f(x)(1-\Phi(\tilde{s}))+q(1-\Psi(\tilde{s})))(C_C-C_M-\pi_M)}{(f(x)\phi(\tilde{s})+q\frac{1-s}{s}\phi(\tilde{s}))(D(T)-\Pi(T))}$. Since $\frac{\beta^T\pi_M}{\Pi^T} = \frac{\beta^T(C_C-C_M-\pi_M)}{D(T)-\Pi(T)}$,this is equivalent to comparing $\frac{1-\Phi(\tilde{s})}{\phi(\tilde{s})}$ with $\frac{f(x)(1-\Phi(\tilde{s}))+q(1-\Psi(\tilde{s}))}{f(x)\phi(\tilde{s})+q\frac{1-s}{s}\phi(\tilde{s})}$.

By the definition of $\Psi(\tilde{s})$, we note that $\frac{\Psi(s)}{\Phi(s)} > \frac{1-s}{s} > \frac{1-\Psi(s)}{1-\Phi s}$ for any $s \in (0, 1)$. Thus $\frac{1-\Psi(\tilde{s})}{\frac{1-s}{s}} < 1 - \Phi(\tilde{s})$ which implies that $\frac{f(x)(1-\Phi(\tilde{s}))+q(1-\Psi(\tilde{s}))}{f(x)\phi(\tilde{s})+q\frac{1-s}{s}\phi(\tilde{s})} < \frac{1-\Phi(\tilde{s})}{\phi(\tilde{s})}$. So $\frac{\beta^T\pi_M(1-\Phi(\tilde{s}))}{\phi(\tilde{s})\Pi(T)} > \frac{\beta^T(f(x)(1-\Phi(\tilde{s}))+q(1-\Psi(\tilde{s})))(C_C-C_M-\pi_M)}{(f(x)\phi(\tilde{s})+q\frac{1-s}{s}\phi(\tilde{s}))(D(T)-\Pi(T))}$ and thus:

$$\frac{\beta^T\pi_M(1-\Phi(\tilde{s}))(f(x)\phi(\tilde{s})+q\frac{1-s}{s}\phi(\tilde{s}))(D(T)-\Pi(T))}{\phi(\tilde{s})\Pi(T)} > \beta^T(f(x)(1-\Phi(\tilde{s}))+q(1-\Psi(\tilde{s})(C_C-\pi_M-C_M) \tag{16}$$

Because $\frac{dW}{d\tilde{s}} = 0$ implies that:

$$\frac{\phi(\tilde{s})f\prime(x)}{-f\prime\prime(x)(1-\Phi(\tilde{s}))^2\Pi(T)}(C(T)+\Phi(\tilde{s})D(T)) = (\phi(\tilde{s})f(x)+q\phi(\tilde{s})\frac{1-s}{s})(D(T)-\Pi(T)) \tag{17}$$

and both sides must be positive since all terms are positive, we can write the numerator of the left side of (2.7) as $\beta^T \pi_M (1-\Phi(\tilde{s}))\frac{f\prime(x)}{-f\prime\prime(x)(1-\Phi(\tilde{s}))^2\Pi(T)}(C(T)+\Phi(\tilde{s})D(T))$. Canceling and combining terms gives:

$$\frac{\beta^T \pi_M f\prime(x)}{-f\prime\prime(x)(1-\Phi(\tilde{s}))\Pi(T)}(C(T)+\Phi(\tilde{s})D(T)) > \beta^T (f(x)(1-\Phi(\tilde{s}))+q(1-\Psi(\tilde{s})))(C_C-(\pi_M+C_M)) \tag{18}$$

and, according to (4), $\frac{dW}{dT} > 0$ ■

Proof. Proof of Lemma 1. ■

Formally,(6)=0 implies that

$$\frac{\phi(\tilde{s})f\prime(x)}{-f\prime\prime(x)(1-\Phi(\tilde{s}))^2\Pi(T)}(C(T)+\Phi(\tilde{s})D(T)) = (f(x)\phi(\tilde{s})+q\psi(\tilde{s}))(D(T)-\Pi(T)) \tag{19}$$

–

We are interested in the sign of this difference that reflects the welfare effect of decreasing the standard of proof for low types:

$$\frac{\phi_L(\tilde{s})f\prime(x_L)}{-f\prime\prime(x_L)(1-\Phi_L(\tilde{s}))^2\Pi(T)}(C(T)+\Phi_L(\tilde{s})D(T)) - (f(x_L)\phi_L(\tilde{s})+q\psi(\tilde{s})(D(T)-\Pi(T)) \tag{20}$$

Let us divide the first equation by $\frac{\phi(\tilde{s})}{1-\Phi(\tilde{s})}$, and let us divide the difference by $\frac{\phi_L(\tilde{s})}{1-\Phi_L(\tilde{s})}$. (19) then becomes

$$\frac{\Pi(T)f\prime(x)}{-f\prime\prime(x)(1-\Phi(\tilde{s}))\Pi(T)^2}(C(T)+\Phi(\tilde{s})D(T)) = (1-\Phi(\tilde{s}))((f(x)+q\frac{1-s}{s})(D(T)-\Pi(T)) \tag{21}$$

while (20) becomes

$$-f\prime\prime(x_L)(1-\Phi_L(\tilde{s}))\Pi(T)(C(T)+\Phi_L(\tilde{s})D(T))-((1-\Phi_L(\tilde{s}))(f(x_L)+q\frac{\phi(\tilde{s})(1-\Phi_L(\tilde{s}))}{\phi_L(\tilde{s})}\frac{1-s}{s})(D(T)-\Pi(T)) \tag{22}$$

–

We know that $f(x_L) < f(x)$, so $f\prime(x_L) > f\prime(x)$ and $(1-\Phi_L(s)) < (1-\Phi(s))$, and also because of the MLRP, we know that $\frac{\phi_L(s)}{(1-\Phi_L(s)} > \frac{\phi(s)}{1-\Phi(s)}$ thus the first term of the difference in (22) is larger than the left side of the equation in (21). Because of MLRP, the second term in (22) is smaller than the right side of (21),so the difference is positive and $\frac{dW}{d\tilde{s}}$ is negative.

Proof. Proof of Lemma 2 ■

By the first order condition on the signal threshold $\tilde{s}$, we know that (changing the profit and welfare terms from aggregate to per period)

$$\pi_M(\lambda S\phi + .(1-\lambda)S_L\phi_L) = (\lambda f(x)\phi + .(1-\lambda)f(x_L)\phi_L + q\psi)(W_C - W_M) \tag{23}$$

If $T < \hat{T}$, then T must satisfy the first order condition on patent length, and

$$\pi_M(\lambda S(1-\Phi)+.(1-\lambda)S_L(1-\Phi_L)) = (\lambda f(x)(1-\Phi)+.(1-\lambda)f(x_L)(1-\Phi_L)+q(1-\Psi))(W_C-W_M) \tag{24}$$

Multiplying (23) by $\frac{(1-\Phi_L)}{\phi_L}$ and subtracting it from (24), we have the effect of increasing patent length while increasing standard of proof just enough to keep the expected rewards from innovation constant for low type inventors. We divide by $\lambda\pi_M$ for compactness, and obtain:

$$S_H((1-\Phi)-(1-\Phi_L)\frac{\phi}{\phi_L}) = (f(x)((1-\Phi)-(1-\Phi_L)\frac{\phi}{\phi_L})+\frac{q}{\lambda}((1-\Psi)-(1-\Phi_L)\frac{\psi}{\phi_L})\frac{W_C-W_M}{\pi_M} \tag{25}$$

Note that by the MLRP, $(1-\Psi) < (1-\Phi_L)\frac{\psi}{\phi_L}$, so

$$S_H((1-\Phi)-(1-\Phi_L)\frac{\phi}{\phi_L}) < (f(x)((1-\Phi)-(1-\Phi_L)\frac{\phi}{\phi_L})\frac{W_C-W_M}{\pi_M}$$

Since the MLRP implies that $(1-\Phi_H) > (1-\Phi_L)\frac{\phi_H}{\phi_L}$, this implies that

$$S_H < f(x)\frac{W_C-W_M}{\pi_M}$$

Proof. Proof of Proposition ?

For brevity, we abbreviate $\Phi(\tilde{s}), \phi(\tilde{s}), \Phi_L(\tilde{s})$ and $\phi_L(\tilde{s})$ as Φ, ϕ, Φ_L , and ϕ_L. Similarly, we use $\hat{\Phi}$ to refer to $\Phi(\hat{s})$ and so on. Note that since we have defined the maximum signal as $s = 1$. the two lever policy (s, T) is equivalent to a three lever policy $(s, \hat{s}, T)$, where $\hat{s} = 1$. Consider any three lever policy $(s, \hat{s}, T)$, such that s and T are set optimally conditional on $\hat{s}$, so that $\frac{dW}{ds} = \frac{dW}{dT} = 0$

By lemma 2,

$$S_H < f(x)\frac{W_C-W_M}{\pi_M}$$

.

Now consider the effects of changing the top threshold, $\hat{s}$.

$$\frac{dW}{d\hat{s}} = \pi_M[\lambda S_H\phi + (1-\lambda)S_L\phi_L] - [\lambda f(x)\phi + (1-\lambda)f(x_L)\hat{\phi}_L + q\hat{\psi})(W_C-W_M) \tag{26}$$

Let us look at the effects of combining a change in $\hat{s}$ with a simultaneous change in patent length that would keep total welfare from low types and obvious inventors constant.

Define $\zeta = \pi_M(1-\lambda)S_L(\hat{\Phi}_L-\Phi_L)-((1-\lambda)f(x_L)(\hat{\Phi}_L-\Phi_L)+q(\hat{\Psi}-\Psi))(W_C-W_M)$ we know this is positive, since $\zeta = \frac{dW}{dT} - (\hat{\Phi}-\Phi)(\pi_M S_H - f(x)(W_C-W_M)$, and $\frac{dW}{dT} = 0$.

Define $\xi = \pi_M \lambda S_L \hat{\phi}_L - (\lambda f(x_L)\hat{\phi}_L + q\hat{\psi})(W_C - W_M)$ We know that this is positive as well, because it is positive at $\tilde{s}$– so more so at $\hat{s}$

Multiply

$$\pi_M(\lambda S_H(\hat{\Phi}-\Phi+(1-\lambda)S_L(\hat{\Phi}_L-\Phi_L))-(\lambda f(x_H)(\hat{\Phi}-\Phi+.(1-\lambda)f(x_L)(\hat{\Phi}_L-\Phi_L)+q(\hat{\Psi}-\Psi))(W_C-W_M) \tag{27}$$

(which is 0) by $\frac{\xi}{\zeta}$ and subtract it from 26. We now have

$$\frac{dW}{d\hat{s}} = \lambda(\pi_M S_H(\hat{\phi} - \frac{\xi}{\zeta}(\hat{\Phi} - \Phi)) - f(x)(\hat{\phi} - \frac{\xi}{\zeta}(\hat{\Phi} - \Phi))(W_C - W_M))$$

Because $S_H < f(x)\frac{W_C - W_M}{\pi_M}$, this will have the opposite sign of $\hat{\Phi} - \frac{\xi}{\zeta}(\hat{\Phi} - \Phi)$. Because $\frac{\hat{\phi}_L}{\hat{\psi}} > \frac{\hat{\Phi}_L - \Phi_L}{\hat{\Psi} - \Psi_L}$, we know that $\frac{\xi}{\zeta} < \frac{\hat{\phi}_L}{\hat{\Phi}_L - \Phi_L}$.

We also know that $\frac{\hat{\phi}}{\hat{\Phi} - \Phi} > \frac{\hat{\phi}_L}{\hat{\Phi}_L - \Phi_L}$, since the MLRP implies that $\frac{\phi(\hat{s})}{\phi_L(\hat{s})} > \frac{\hat{\Phi} - \Phi(s)}{\hat{\Phi}_L - \Phi_L(s)}$ for any $s < \hat{s}$, so $\frac{\xi}{\zeta} < \frac{\hat{\phi}}{\hat{\Phi} - \Phi}$ Thus $\hat{\phi} > \frac{\xi}{\zeta}(\hat{\Phi} - \Phi)$, so

$$\frac{dW}{d\hat{s}} = (1 - \lambda)(\pi_M S_H(\hat{\phi} - \frac{\xi}{\zeta}(\hat{\Phi} - \Phi)) - f(x)(\hat{\phi} - \frac{\xi}{\zeta}(\hat{\Phi} - \Phi))(W_C - W_M)) < 0$$

Since $\frac{dW}{d\hat{s}} < 0$, social welfare can be increased by decreasing $\hat{s}$, so any triple $(\tilde{s}, \hat{s}, T)$ where $T < \hat{T}$ cannot be optimal.

■

Proof. Proof of Proposition ?? ■

Recall that we are assuming that $\pi_N = \pi_I$ so (11) becomes:

$$S\theta(\pi_M - \pi_I)/f(x) = y(W_I - W_M) + (1 - y)(W_N - W_M)$$

or

$$S\theta(\pi_M - \pi_I)/f(x) = (W_N - W_M) + y(W_I - W_N)$$

i.e.

$$S\theta(\pi_M - \pi_I)/f(x) = (W_N - W_M) - y(W_N - W_I)$$

solving for y

$$y = \frac{(W_N - W_M) - S\theta(\pi_M - \pi_I)/f(x)}{W_N - W_I}$$

By the definition of ρ, $\pi_M - \pi_I = \pi_M - \pi_N = \rho\pi_M \frac{W_N - W_M}{W_C - W_M}$, so

$$y = \frac{(W_N - W_M) - S\frac{\theta}{f(x)}\rho\pi_M \frac{W_N - W_M}{W_C - W_M}}{W_N - W_I}$$

Recall that $S = \frac{f(x)(W_C - W_M)}{z\pi_M}$, so $y = \frac{(W_N - W_M) - \frac{\theta\rho}{z}(W_N - W_M)}{W_N - W_I}$

Proof. Proof of Comparative Statics Proposition

Using our definition of ρ, we can write $y = \frac{z(W_N - W_M)\pi_M - \theta(W_C - W_M)(\pi_M - \pi_N)}{z(W_N - W_I)\pi_M}$. Inspection reveals that whenever $y > 0$, y is increasing in W_I. Differentiating y with respect to W_N gives:

$$\frac{dy}{dW_N} = \frac{-z(W_I - W_M)\pi_M + \theta(W_C - W_M)(\pi_M - \pi_N)}{z(W_N - W_I)^2\pi_M} \tag{28}$$

Our condition $\frac{\pi_N}{W_C - W_N} > \frac{\pi_M}{W_C - W_M} > \frac{\pi_I}{W_C - W_I}$ that defines what is an infringing versus a non-infringing product implies that $(W_C - W_M)(\pi_M - \pi_N) > (W_I - W_M)\pi_M$ if $\pi_N = \pi_I$, as we have assumed. Thus, since $\theta > z$, this is positive. Differentiating y with respect to W_M gives:

$$\frac{dy}{dW_M} = \frac{-z\pi_M + \theta(\pi_M - \pi_N)}{z(W_N - W_I)\pi_M} \tag{29}$$

■

This is positive if and only if $z/\theta < (\pi_M - \pi_N)/\pi_M$. Differentiating y with respect to π_M gives:

$$\frac{dy}{d\pi_M} = -\frac{\theta(W_C - W_M)\pi_N}{z(W_N - W_I)\pi_M^2} < 0 \tag{30}$$

Differentiating y with respect to π_N gives:

$$\frac{dy}{d\pi_N} = \frac{\theta(W_C - W_M)}{z(W_N - W_I)\pi_M} > 0 \tag{31}$$

7 References

Denicolo, Vincenzo. 1996. “Patent Races and Optimal Patent Breadth and Length" 44 *The Journal of Industrial Economics* 249-265.

Gilbert, R and Carl Shapiro, 1990 “Optimal Patent Length and Breadth” 21 *Rand Journal of Economics*, 106-112.

Holmstrom, Bengt and Paul Milgrom, 1987, ?Aggregation and Linearity in the Provision of Intertemporal Incentives,? *Econometrica*, 55 (2), 303-328.

Klemperer, Paul,1990, 'How Broad Should the Scope of Patent Protection Be?', 21 *Rand Journal of Economics*, 113-130.

MIrrlees, , James A. ,1974, ?Notes on Welfare Economics, Information and Uncertainty,? in M. Balch, D. McFadden, and S. Wu, eds., *Essays in Equilibrium Behavior under Uncertainty*, Amsterdam: North-Holland,

Yelderman, Stephen, 2017, "The Value of Accuracy in the Patent System", *Univ of Chicago Law R.* 84:1217