On the Interaction between Patent Screening and its Enforcement*

Gerard Llobet
CEMFI and CEPR

Álvaro Parra
UBC Sauder

Javier Suarez
CEMFI and CEPR

October 6, 2023

Abstract

Using an industry-dynamics model of innovation, we explore the interplay between patent screening and patent enforcement. Costly enforcement involves type I and type II errors. When the patent office takes the rates at which such errors occur as given, granting some invalid patents is socially optimal even with costless screening because it encourages innovation. When the influence on courts' enforcement efforts is considered, these same forces imply that screening and enforcement are complementary. That is, contrary to common wisdom, better screening induces better enforcement, and an increase in enforcement costs could be optimally accommodated with less rather than more ex-ante screening.

JEL Codes: L26, O31, O34.
Keywords: Intellectual Property, Innovation, Imitation, Patent Screening, Patent Enforcement, Industry Dynamics.

*This work is based on a previous paper entitled "Entrepreneurship Innovation, Patent Protection, and Industry Dynamics." Financial support from the European Commission (grant CIT5-CT-2006-028942) and the Spanish Ministry of Science and Innovation through the ECO2008-00801, the Consolider-Ingenio 2010 Project "Consolidating Economics" and the Regional Government of Madrid through grant S2015/HUM-3491 is gratefully acknowledged. Address for correspondence: Gerard Llobet, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain. Tel: +34-914290551. Fax: +34-914291056. Email: llobet@cemfi.es.

1 Introduction

In recent decades, we have witnessed a surge in patenting activity. The large number of applications has put a strain on patent offices everywhere. There is a concern that this process might have led to a proliferation of likely invalid patents that could hinder future technological progress, particularly in areas where innovation is cumulative. Between 75% and 97% of the applications reviewed by the US Patent and Trademark Office (USPTO) end up in a patent being granted (Lemley and Sampat, 2008). To address this issue, many authors advocate increasing the resources allocated to patent offices (c.f., Farrell and Merges, 2004).[1] Others, however, argue that the high approval rates are the consequence of *rational ignorance* (Lemley, 2001): Since only a tiny fraction of patents is ever litigated,[2] performing in-depth ex-post screening through litigation, rather than carefully analyzing ex-ante whether each patent is valid and relevant, can be more cost-effective.[3]

In this article, we study the interplay between ex-ante screening by the patent office and ex-post enforcement by courts and their impact on innovation and welfare. We show that when courts are imperfect—i.e., there is a chance to make an incorrect ruling—and these mistakes are regarded independent from the patent office's behavior, some rational ignorance by the patent office is socially optimal even in the *absence* of screening costs. That is, even when ex-ante screening could be perfect at no cost, allowing a percentage of invalid patents could be optimal. However, when courts' mistakes depend on judges' endogenous effort (as in Daughety and Reinganum, 2000) ex-ante screening and ex-post enforcement are complementary, that is, an increase in screening by the patent office induces higher enforcement effort by courts, leading to better rulings.

Understanding the influence of the patent office's screening behavior on the courts' enforcement decisions requires a dynamic approach. From a static perspective, there is no

[1]This view has been adopted in the Leahy-Smith America Invents Act of 2011 that increases funding and provides new screening mechanisms to discern the quality of patent applications better.

[2]Lemley (2001) estimates that 2% of patents are litigated and only about 0.2% of patents reach a courtroom.

[3]Frakes and Wasserman (2019) challenge the results of this cost-benefit analysis.

interaction; judges make their decisions after innovation happens, generating no downsides to allowing entry. From a dynamic perspective, however, judges' decisions determine the market's competitive state, affecting rents and the prospects of future entry. Screening affects the entry quality of innovators, influencing the costs of the judge's mistakes. To study this interaction, we propose a tractable industry-dynamics model with sequential innovation and endogenous entry. The industry consists of a continuum of business niches, each of which can be thought of as the market for a distinct product. Successful developers of improved versions of each product contribute to welfare and appropriate temporary monopoly profits as in a standard quality ladder model with limit pricing (Grossman and Helpman, 1991; Aghion and Howitt, 1992). These temporary monopolies are based on intellectual property rights (IPRs) and are threatened by the endogenous arrival of two kinds of entrants: developers of better versions of the product (that we denote as genuine innovators) and entrants that contribute minor variations with little social value (that we denote as obvious innovators).[4]

In every period, entrants observe market conditions and invest in R&D until the quasi-rents from entry are dissipated. Entrants face uncertainty on whether their product will be opposed on the basis of the validity of its patent or the infringement of the incumbent's patent rights and, as a result, they suffer from the "tragedy of anticommons" (Heller and Eisenberg, 1998). This assumption is consistent with Lemley (2008), who argues that, due to the large number of overlapping rights, firms decide to innovate first and deal later with the lawsuits that ensue from existent patent holders. This strategy is also supported by the large proportion of patents brought to court that end up invalidated (Allison and Lemley, 1998).

Upon entry, firms discover whether their product constitutes a genuine or an obvious innovation. Entrants apply first for a patent and only learn the quality of their innovation through the commercialization of their products. Patent applications are presumed valid,

[4]We assume that obvious innovations do not fulfill the novelty requirements for patentability nor represent a sufficient innovative step to place them outside the breadth of existing patents. Genuine innovations satisfy both.

as patent examiners have to find prior art and articulate an appropriate basis for rejection. This means that genuine innovators always receive a patent, but the success of an obvious innovator depends negatively on the amount of resources that the patent office devotes to screening—the screening rate. This assumption is consistent with the findings in Frakes and Wasserman (2017), which shows that a lack of resources for patent examiners (e.g., tighter time constraints) results in an upward bias in the approval rate.[5]

After a patent has been obtained, an entrant may randomly reach a competitive niche or one monopolized by an incumbent. A genuine innovator monopolizes a competitive niche, while an obvious innovator keeps the niche competitive as it introduces a product of similar quality to those already in the market. If the entrant lands in a monopolized niche, the incumbent goes to court to preserve its rents by claiming that the entrant infringes on its patent. If the court rules in favor of the entrant, both firms compete in the same niche, driving the incumbent's profits to zero, whereas the entrant makes profits according to its quality. If the court determines that the entrant has infringed the incumbent's patent, the innovation goes to waste.

The strength of patent protection is endogenous and involves dynamic considerations absent in the pre-existing literature. Each case that arrives to court is decided by a different judge. Judges make evidence-based decisions; they can err in their rulings because this evidence can be noisy or misinterpreted. Consequently, patent protection is probabilistic (c.f., Lemley and Shapiro, 2005; Farrell and Shapiro, 2008).[6] For each case, a judge decides how much costly effort to devote to analyzing the evidence. Although both the patent office and judges choose their efforts to maximize social welfare, they make their decisions at different stages of the entry process. The patent office oversees every patent application, while courts only evaluate the validity of a patent conditional on an

[5]The authors also show that the bias towards patenting created by shortening the allocated time for the reviewing process is more prominent in industries where technologies are complex and innovation is sequential, as in the framework discussed in this paper. See also Lei and Wright (2017).

[6]Other works in which courts make probabilistic rulings include Spier (1994), Daughety and Reinganum (1995), and Landeo et al. (2007). By adopting this probabilistic approach, we abstract away from the traditional patent length and breadth discussion (Scotchmer, 2004).

entrant having reached a monopolized niche. As explained below, this asymmetry creates a dynamic inconsistency problem within the patent system. That is, judges will see some of their errors as a mechanism to foster competition and, through it, the entry of future genuine innovators.

A judge's objective function can be written as a weighted average of the welfare costs of committing type I and type II errors. A type I error arises when a judge rules against an entrant with a genuine innovation, depriving society of that improvement's benefit. A type II error arises when an obvious innovator is allowed to compete with the existing incumbent. Type II errors have non-trivial effects on entry and welfare. On the one hand, they shorten the expected duration of the incumbency of genuine innovators, discouraging entry. On the other hand, type II errors modify market structure: they turn previously monopolized niches competitive, producing two kinds of benefits. First, the social costs of the existing monopoly are dissipated. Second, the entrants' prospect of facing opposition in a niche improves, encouraging entry. That is, allowing obvious innovators to compete mitigates the distortions associated with the tragedy of the anticommons. On the net, in our model, type II errors always have a strictly positive effect on entry and welfare. This finding is consistent with Galasso and Schankerman (2015), who show that patent invalidation positively correlates with future entry.

Despite the benefits of incurring a type II error, the joint social costs of both errors are always positive. Better screening by the patent office increases the importance of type I errors relative to type II errors and fosters judges' effort. This complementarity between ex-ante screening and ex-post enforcement is further reinforced by the complementarity between the decisions of the current and future judges in a given niche. When the rulings of the judges that will oversee the same niche in the future become more accurate, genuine innovators are more likely to succeed regardless of the competitive state of the niche. As a result, the gains from altering market structure that result from the type II error are reduced, increasing the current cost of an inaccurate ruling.

The impact of patent screening on the costs balanced by the judges when considering the implications of their enforcement errors has implications for the socially-optimal level of patent screening taking into account its impact on the endogenous response of ex-post enforcement. The optimal level of patent screening balances off two forces. On the one hand, there is the traditional substitution effect consistent with the idea of rational ignorance. A decrease in enforcement costs, which leads to more enforcement, should be accommodated with less ex-ante screening to decrease the overall costs of the patent system. On the other hand, and due to the complementarities described above, less ex-ante screening also induces worse enforcement. The presence of this second effect implies that relative to a situation in which enforcement is exogenous, socially optimal screening is higher when the courts' endogenous response is taken into account.

Our key findings are robust to variations of our main assumptions. Section 5 includes several extensions showing that the uncovered trade-offs and the strategic complementarity between screening and enforcement still hold when we consider cost-saving innovations and the possibility of innovation generating deadweight losses; continuous, rather than binary, screening effort by judges; judges operating under biased priors favoring the entrant or the incumbent; situations in which obvious incumbents hold-up of future entrants extracting rents via licensing (i.e., patent trolling), and; when the knowledge from genuine innovations precluded to enter the market is not wasted, but incorporated in the public domain so that future innovators build upon that technology.

The article is organized as follows. We next discuss the related literature. Section 2 introduces the baseline model. Section 3 shows that, for a given time-invariant combination of screening intensity decided by the patent office and enforcement intensity decided by judges, the model displays a unique steady-state equilibrium. We provide the comparative statics of such equilibrium and characterize the socially optimal intensities of screening and enforcement in a frictionless world where a social planner could costlessly set both. In Section 4, we endogenize ex-ante screening and ex-post enforcement, identifying the dynamic trade-offs that constitute the central contribution of the paper. Section 5 discusses extensions of our framework, and Section 6 concludes. All proofs are in the Appendix.

Related Literature To our knowledge, this is the first article providing a formal framework to understand the interaction between patent screening and enforcement and their impact on innovation and market dynamics. Our analysis builds upon several strands of literature.

In our model, innovation is sequential (or cumulative). In that strand of the literature, various dimensions of patent policy have already been studied, including patentability requirements (Scotchmer and Green, 1990; O'Donoghue, 1998), patent breadth and length (O'Donoghue et al., 1998), forward protection (Denicolò, 2000; Denicolò and Zanchettin, 2002), or lack of protection (Bessen and Maskin, 2009). Other works within the sequential innovation framework study its antitrust implications (Segal and Whinston, 2007), optimal buyouts schemes (Hopenhayn et al., 2006), growth and industry dynamics (Denicolò and Zanchettin, 2014), and product-market competition (Marshall and Parra, 2019). We add to this literature by exploring the dynamic trade-offs that shape the interaction between patent screening and enforcement.

Our contribution adds to the prior literature on entry efficiency and IPRs. Dixit and Stiglitz (1977) show that entry is insufficient when entrants cannot appropriate all the new rents they generate. Patent screening and enforcement limit appropriability in two ways: affecting the duration of the incumbency status (as in O'Donoghue et al., 1998) and acting as an entry barrier (Gilbert and Newbery, 1982). We show that these mechanisms are in tension, making imperfect patent screening desirable. In this respect, we complement the analysis of Parra (2019), which studies the optimal patent design when market structure is endogenously determined by (an exogenous) patent strength. In our setup, patent strength is an endogenous object determined by the quality of screening by the patent office, the intensity of subsequent entry, and the enforcement effort of the courts.

The law and economics literature has long recognized that courts might be imperfect in their rulings (see Spier, 2007, for an extensive survey on litigation). When endogenizing courts' decisions, the literature generally assumes that prosecutors are driven by social welfare considerations (Grossman and Katz, 1983; Reinganum, 1988), career concerns (Daughety and Reinganum, 2020), or both (Daughety and Reinganum, 2016). We consider welfare-maximizing judges who can improve the quality of their rulings (reduce their errors) by exerting costly effort.

Finally, our paper contributes to the literature that endogenizes the effective strength of patent rights. Caillaud and Duchêne (2011) study how congestion at the patent office leads to poorer screening, inducing low-quality innovators to apply for a patent. Atal and Bar (2014) explore the equilibrium consequences on R&D, patent quality, and welfare of introducing a two-tier patent system discriminating on innovation quality. Schankerman and Schuett (2021) study various policy instruments, such as screening intensity and screening fees, in a framework where patent challenges emerge endogenously, and court rulings are perfect. Building upon this literature, we allow the enforcement effort by courts to respond endogenously to the patent office's screening quality. This way, we can unveil the potential complementarity between screening and enforcement.

Picard and van Pottelsberghe de la Potterie (2013) also study the relationship between screening and enforcement. In their model, judges make probabilistic rulings based on their posterior belief that entry is genuine. Screening, thus, complements enforcement by improving the judges' decisions via refined posterior beliefs. In our model, enforcement is an endogenous strategic decision, and screening may act as its strategic complement or substitute. Screening complements enforcement by reducing the probability of obvious entry, increasing the expected (dynamic) cost of leaving an entrant out of the market, and incentivizing judges to make accurate rulings. Screening also substitutes enforcement by decreasing the total number of cases reaching court, decreasing the patent system's cost.

2 The Baseline Model

We characterize the evolution of an industry in an infinite-horizon discrete-time model with discount factor $\beta < 1$. This industry is comprised of a continuum of business niches of measure one. Each niche can be interpreted as the market for a different product.[7] Within a niche, firms compete in prices and through innovations in the context of a quality ladder. The niche's demand is unit-inelastic and supported by a unit mass of infinitely-lived homogeneous consumers. At each date t, consumers derive a net utility flow $U_t = Q_t - P_t$ from consuming a good of a niche with quality Q_t at a price P_t. Utility is additive across goods and dates, and production costs are assumed to be zero.

2.1 Untargeted Entry and Patent Screening

In every period t, there is an endogenous measure $e_t \in [0, 1]$ of entrants extracted from a large population of potential firms that face an entry cost normalized to 1. Each entrant develops a new product (innovation) and applies for a patent. Each innovation can be genuine or obvious with exogenous probabilities α and $1 - \alpha$, respectively. Depending on the quality of the last introduced product, niches can be monopolized or competitive. We denote the proportion of monopolized niches as $x_t \in [0, 1]$.

A genuine innovation improves the quality of the existing technology by $\pi > 0$. When allowed to reach the market, a genuine entrant replaces the existing incumbent, monopolizing the niche. Price competition and the unit-demand assumption imply that a monopolistic incumbent earns a per-period profit flow of π.[8] An obvious innovation, in turn, provides negligible improvements, becoming a perfect substitute for the existing product. If allowed to produce, an obvious entrant competes with the existing incumbent, turning the niche competitive and driving profits to zero.

[7]This simplification allows us to abstract from cross-product competition to focus on competition related to concomitant and future entry into each niche.

[8]In equilibrium, genuine entrants charge a price $P_t = \pi$, capturing as profit the full quality improvement and generating no deadweight loss. Section 5.1 discusses an alternative environment where firms invest in cost-reducing innovations, and consumers' demand is not unit-elastic. In that case, the exercise of market power in monopolized niches implies a deadweight loss, altering the welfare implications.

A *patent office* screens all applications and determines whether to grant a patent. Applications based on genuine innovations are always successful, whereas those based on obvious innovations succeed with probability λ. We interpret $1 - \lambda$ as a measure of the patent office's screening quality.[9] Only innovators that receive a patent can enter a niche. Potential entrants decide on entry without knowing their innovations' quality, which they learn by competing in their niche. Entry is untargeted and uniformly distributed over the existing niches so that, in every niche and period, there will be an independent-across-niches probability e_t of having just one entrant and a probability $1 - e_t$ of having no entrant. The probability that an entrant lands in a monopolized niche is x_t, which is independent of the entrants' type. Entrants' uncertainty regarding the niche they will occupy captures the "tragedy of anticommons," which states that, due to the many overlapping rights, firms decide to innovate first and deal later with the lawsuits from existent patent holders.[10]

2.2 Patent Enforcement

Incumbent firms in monopolized niches might lose their status due to either competition with an obvious innovator (i.e., a firm with a product of similar quality) or to the replacement by a genuine innovator (a firm with a superior substitute product). An incumbent can respond to entry by filing a lawsuit to prevent the entrant from producing. For simplicity, we assume that litigation is costless, but the incumbent avoids it whenever indifferent. This assumption means incumbents in already competitive niches (i.e., with no profits at stake) will not engage in litigation.

Each case is reviewed by a different judge who makes a probabilistic decision. If the judge rules in favor of the incumbent, entry is blocked, allowing the incumbent to preserve

[9] Because applications are presumed valid, our modeling can be interpreted as the result of a search for prior art. Thus, the parameter λ represents the probability that the patent office fails to find similar existing products when they exist, deeming the innovation genuine.

[10] This formulation also simplifies exposition by abstracting away from competition between simultaneous innovators, which would introduce niche congestion—as in the literature on random search—and patent races (e.g., Loury, 1979; Lee and Wilde, 1980).

the monopoly status. To simplify the exposition, we assume the entrant's innovation goes to waste when entry is blocked.[11] If the judge rules in favor of the entrant, this firm replaces the existing incumbent and receives a profit flow according to the quality of its improvement. A judge ruling favors an entrant with an obvious and a genuine innovation with endogenous probabilities denoted by μ_0 and μ_1, respectively.

Notice that the judge implicitly decides whether the entrant infringes the incumbent's patent *and* whether the entrant's patent is valid. Due to the nature of the innovation, these two decisions boil down to assessing the size of the inventive step. When the innovation is genuine, the quality improvement π is assumed to be sufficient to deserve a patent and, simultaneously, to be outside the breadth of existing IPRs. When the innovation is obvious, the new product is assumed not to satisfy the non-obviousness requirement necessary for being patentable and, consequently, to infringe on the incumbent's patent.[12]

We assume that judges make evidence-based rulings. They rule in favor of the incumbent if and only if their fair interpretation of the evidence allows them to conclude that the entrant's innovation is obvious. When a case reaches the court, the judge, who does not directly observe the quality of the entrant's innovation, can exert effort (that is, invest resources) to receive a costly signal σ about the merit of the incumbent's case. The outcome of the signal is binary, taking a value 0 when the judge concludes that the innovation is obvious and a value of 1 otherwise. The precision of this signal depends on the unobservable evidence-gathering effort of each judge, $s \in [0, 1]$, according to the following simple specification:

$$\mu_0(s) = \Pr[\sigma = 0|\text{obvious}] = \frac{1-s}{2} \quad \text{and} \quad \mu_1(s) = \Pr[\sigma = 0|\text{genuine}] = \frac{1+s}{2}. \tag{1}$$

Thus, if no effort is exerted, $s = 0$, the signal classifies genuine and obvious innovators as obvious with equal probability, $\mu_0(0) = \mu_1(0) = 1/2$. Under maximum effort, $s = 1$, the signal perfectly separates genuine and obvious innovations, $\mu_0(1) = 0$ and $\mu_1(1) = 1$.

[11] Section 5.4 shows that the results of our model extend when future entrants can build on innovations of unsuccessful entrants.

[12] Because entry drives the incumbent profits to zero and firms only litigate to defend their profits, ruling in favor of the entrant is also equivalent to the *de facto* invalidation of the incumbent's patent.

Judges face a cost of effort $c(s)$ increasing in s. This cost captures the effort of gathering evidence, analyzing, and deliberating on the case.[13]

Importantly, we assume each case is overseen by a different independent judge. This judge decides how much effort to exert to maximize the social surplus (welfare) associated with determining the right of the entrant to produce in the niche under dispute. As we explain later, this welfare maximization is akin to minimizing the weighted cost of type I errors (not allowing a genuine innovation to be implemented) and type II errors (allowing an entrant with an obvious innovation to compete with the incumbent). In doing so, each judge takes the effort of the other judges as given, as well as the screening rate of the patent office, λ.

2.3 The Effects of Licensing

Our model abstracts from the existence of licensing. In a frictionless environment, licensing could circumvent the distortions uncovered in this paper. Consider, for example, a situation where an incumbent monopolist licenses its patent to a genuine entrant while, at the same time, it withdraws its product from the market. Absent competition, the contribution of the incumbent's product would no longer be transferred to the consumers through a lower price. As a result, the per-period rents of the entrant, 2π, exceed the total profits that both firms obtain in our benchmark model, π. This increase in profits would be enough to support a licensing fee that makes the incumbent better off compared to the case of litigation. Another possible arrangement would entail the entrant selling its patent to the incumbent, who would then monopolize the two highest qualities. In both cases, litigation would be averted. At the same time, the rents from innovation would increase, regardless of the value of μ_1.[14]

[13] Section 5.3 explores the implications of assuming that judges have a more utilitarian pro-competitive bias or some pro-incumbent anti-competitive bias. We show that the main insights from the baseline model remain essentially unchanged.

[14] Note that this process is compounded when multiple genuine innovators arrive in sequence. In the previous example, once another genuine arrives, licensing would imply that the two highest qualities below the frontier are no longer sold, and the market price and profits become 3π. This process would continue until an obvious entrant is successful in court and competes these profits away.

Of course, this result is not specific to our model, and it has been pointed out in the literature, along with reasons why frictionless licensing is unrealistic in practice. One of the main reasons is the existence of hold-up (Hopenhayn et al., 2006; Lemley and Shapiro, 2007). When negotiation occurs after investments have taken place and the incumbent has bargaining power, a genuine entrant would anticipate that some of its rents would be expropriated, reducing the incentives to innovate in the first place. The existence of private information (Llobet, 2003), different expectations about the probability that the patents are valid (Choi, 1998), and other transaction costs would also reduce the incentives for firms to reach a licensing agreement and explain why litigation often emerges in practice.

More generally, the licensing strategies outlined above rely on the incumbent's product being withdrawn from the market or offered at a high price, making it *de facto* irrelevant. These kinds of arrangements are likely to be frowned upon by competition authorities as they could be interpreted as collusive and, despite the increase in innovation they entail, they would be harmful to consumers (Green and Scotchmer, 1995).

The previous arrangements (including cross-licensing deals) cannot be sustained once we restrict ourselves to situations where the incumbent and the entrant must compete. Total per-period rents remain π regardless of whether the incumbent or the entrant wins in court, so there is no room for negotiating a licensing agreement, as this paper assumes.

In fact, licensing can even exacerbate the previous distortions on innovation. In Section 5.5, we develop an extension of the model where the incumbent in a competitive niche—i.e., resulting from an obvious innovation—can sue a genuine entrant for part of its rents. In contrast with the previous case, negotiation here is profitable as a licensing agreement guarantees that the per-period rents increase from 0 to π. In contrast, litigation might prevent the entrant from producing with probability $1 - \mu_1$. In this case, the incumbent, a *patent troll*, can extract a share of the future rents via a licensing agreement. By shifting rents from genuine to obvious innovators, we show that, although the incentives to innovate are reduced, the effects uncovered in the next sections are preserved.

3 Exogenous Courts

To ease the exposition and highlight the direct impact of the patent office on equilibrium outcomes, we first solve the model under exogenously given values of the probabilities μ_0 and μ_1, which we will endogenize in Section 4.

For given μ_0 and μ_1, we denote as v_t the value of being the incumbent in a monopolized niche at date t (that is, holding a patent that has not been infringed or whose infringement has been fended off in court) and we can characterize it recursively as

$$v_t = \pi + \beta\{1 - e_{t+1}[(1-\alpha)\lambda\mu_0 + \alpha\mu_1]\}v_{t+1}. \tag{2}$$

This value is composed of the current flow of monopoly profits π and the discounted future value of preserving this position, βv_{t+1}, weighted by the probability of surmounting the potential entry of an innovator at $t+1$. The terms in square brackets take into account that entry occurs with probability e_{t+1}, involves an obvious or a genuine innovator with probabilities $1-\alpha$ and α, respectively, and the probabilities $\lambda\mu_0$ and μ_1 with which each entrant obtains both a favorable assessment by the patent office and a positive court ruling.

As a result of the entry flow and the competition that it might entail, we can write the law of motion governing the proportion of monopolized niches, x_t, as

$$x_{t+1} = x_t[1 - e_{t+1}(1-\alpha)\lambda\mu_0] + (1-x_t)\alpha e_{t+1}. \tag{3}$$

Monopolized niches at $t+1$ are those already monopolized at t that do not experience the successful entry of obvious innovators, $x_t[1 - e_{t+1}(1-\alpha)\lambda\mu_0]$, as such niches become competitive plus the previously competitive niches that transition to a monopoly due to the entry of genuine innovators, $(1-x_t)\alpha e_{t+1}$.

The flow of innovating firms e_t is determined by a free-entry condition. Attempting entry has a cost that we normalize to 1. Entry is profitable, inducing a payoff v_t, only if the entrant obtains a genuine innovation and successfully enters its niche. We denote the probability of a profitable entry by p_t and is given by

$$p_t = \alpha(1 - x_t + x_t\mu_1). \tag{4}$$

An entrant produces a genuine innovation with probability α and successfully enters either because it lands in a competitive niche, which occurs with probability $1 - x_t$, or because it lands in a monopolized niche and surmounts the incumbent's opposition in court, which occurs with probability $x_t\mu_1$. The free-entry condition, thus, can be written as $-1 + p_t v_t \leq 0$, which in an equilibrium involving an interior entry flow $e_t \in (0, 1)$ in period t will hold with equality,

$$p_t v_t = 1. \tag{5}$$

3.1 Steady-State Equilibrium Analysis

Our analysis focuses on the interior-entry equilibrium of the model so that equations (2)-(5) characterize the dynamic equilibrium of the industry under exogenously given values of μ_0 and μ_1. They determine four key endogenous variables at each date t: the proportion of monopolized niches, x_t; the probability of a profitable entry, p_t; the entry flow, e_t; and, the value of being a monopolist, v_t. To ease notation, we will denote the corresponding steady-state value of the above variables simply by x, p, e, and v, respectively.

The following assumption restricts the profit parameter π so that the steady-state equilibrium of the model involves an interior entry flow $e \in (0, 1)$. Lemma 1 shows the necessity and sufficiency of the restriction and provides close-form expressions for the steady-state value of the key variables of such an equilibrium.

Assumption 1. $\pi \in \left(\underline{\pi}, \underline{\pi} + \frac{\beta[\alpha+(1-\alpha)\lambda\mu_0]}{\alpha}\right)$, *where* $\underline{\pi} = \frac{(1-\beta)(\alpha+(1-\alpha)\lambda\mu_0)}{\alpha(\alpha\mu_1+(1-\alpha)\lambda\mu_0)}$.

Lemma 1. *The model has a unique steady-state equilibrium with* $e \in (0, 1)$ *if and only if Assumption 1 holds. This equilibrium is given by*

$$x = \frac{\alpha}{\alpha + (1-\alpha)\lambda\mu_0}, \tag{6}$$

$$p = \alpha\frac{\alpha\mu_1 + (1-\alpha)\lambda\mu_0}{\alpha + (1-\alpha)\lambda\mu_0}, \tag{7}$$

$$e = \frac{\pi p - (1-\beta)}{\beta\left[\alpha\mu_1 + (1-\alpha)\lambda\mu_0\right]}, \tag{8}$$

$$v = \frac{1}{p}. \tag{9}$$

In the above equilibrium, entry occurs until the expected value of developing an innovation, pv, equals the unit entry cost. It is worth noticing that the proportion of monopolized niches, x, and the value of incumbency in steady state, v, are not affected by the profit parameter, π. That is, an increase in monopoly profits is completely offset by increased entry, which raises producers' turnover within the invariant fraction of monopolized niches (reducing the duration of incumbency), allowing the values of x and v to remain unchanged. Notice also that parameters λ and μ_0 always appear in combination, as $\lambda\mu_0$, which represents the rate at which obvious innovations succeed in entering a niche. The next proposition summarizes the comparative statics of this equilibrium.

Lemma 2. *The effect of marginal changes in the parameters on the steady-state equilibrium values of x, p, e, and v have the signs shown in the following table:*

	π	β	α	$\lambda\mu_0$	μ_1
x	0	0	$+$	$-$	0
p	0	0	$+$	$+$	$+$
e	$+$	$+$	$+$	?	$+$
v	0	0	$-$	$-$	$-$

The proportion of niches operating under monopoly, x, is increasing in the probability that an innovation is genuine, α, and decreasing in the rate at which obvious innovations succeed in entering, $\lambda\mu_0$. Intuitively, the higher the probability that a firm with an obvious patent arises and is allowed to produce, $(1\alpha)\lambda\mu_0$, the more often monopolist incumbents will be challenged and defeated in court. In contrast, the probability with which genuine innovators succeed in court vis-a-vis an incumbent patentholder, μ_1, does not affect x since ruling in favor of that entrant implies replacing one monopolist with another.

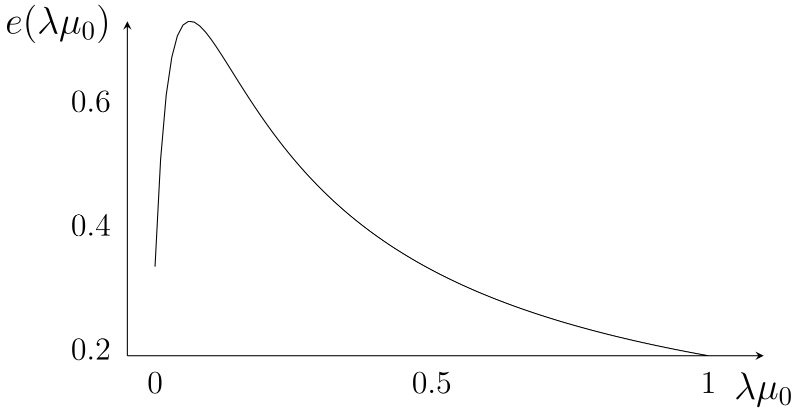
Due to the free-entry condition, the value of incumbency v is inversely related to the probability with which entrants become successful incumbents, p. Such probability increases in the probability that the innovation is genuine, α, and that courts rule in its favor when confronting an incumbent, μ_1. More surprisingly, p is also increasing in the rate at which obvious innovations (which do not directly give rise to incumbency) enter successfully, $\lambda\mu_0$. This effect occurs because the entry of obvious innovators decreases the proportion of monopolized niches x, reducing the probability that a future entrant with a genuine innovator faces the opposition of an incumbent monopolist. This force will make entry not necessarily decreasing in $\lambda\mu_0$ as we will show below.

As expected, entry is increasing in the flow of profits, π, and the discount factor, β. An increase in the judges' probability of ruling in favor of a genuine innovator, μ_1, or in the probability of obtaining a genuine innovation, α, also fosters entry, as it increases the probability of being successful. However, the effect of $\lambda\mu_0$ on entry is generally ambiguous, as the next proposition characterizes.

Lemma 3. *The relationship between the rate at which obvious innovators successfully enter the market, $\lambda\mu_0$, and the steady-state equilibrium entry flow, e, can be increasing, decreasing, or inverted-U shaped. In particular, it is decreasing when $\mu_1 = 1$.*

This result uncovers an interesting non-monotonic relationship between entry and the incumbent's protection against obvious innovations. The main driver of this result is that a change in $\lambda\mu_0$ engenders two effects of opposite sign. On the one hand, an increase in $\lambda\mu_0$ fosters entry — through the decrease in x — as it reduces the proportion of niches in which genuine innovators are challenged in court. On the other hand, an increase in $\lambda\mu_0$ decreases the value of incumbency, v, as monopolists are more likely to see their rents competed away by an entrant with an obvious innovation.

When the probability of success in court of an entrant with a genuine innovation, μ_1, is close to one, the second effect dominates, and entry monotonically decreases with $\lambda\mu_0$. Intuitively, with $\mu_1 = 1$, the innovation-enhancing pro-competitive effect disappears since genuine innovators can always produce, regardless of whether they land in monopolized or competitive niches. As illustrated by Figure 1, however, when the entry of genuine



Note: Parameter values are $\alpha = 0.1$, $\pi = 3.6$, $\beta = 0.8$, and $\mu_1 = 0.6$.

Figure 1: Entry flow and the protection against obvious innovators. This figure shows a case in which entry is maximized at an interior value of the probability with which obvious innovators are allowed to enter, $\lambda\mu_0$.

innovators is not guaranteed, the pro-competitive effect is relevant and may dominate when $\lambda\mu_0$ is low. In those cases, the innovation flow is maximized at some interior value $\lambda\mu_0$. This result will have non-trivial implications for the discussion on the socially optimal level of protection against imitation, $1 - \lambda\mu_0$, and its link to the socially optimal level of protection against a genuine innovation, $1 - \mu_1$.

3.2 Optimal Patent Screening

To gain intuition about the effects of changing the patent office's screening rate λ, we characterize its socially optimal level. Screening affects welfare through the entry rate, the rate at which obvious innovations successfully enter the market, and through the screening costs. Let $\Pi \equiv \pi/(1-\beta)$ represent the social present value generated by a genuine innovation conditional on successful entry. Then, the total per-period net addition to welfare in steady state can be expressed as

$$W = e\big[p\Pi - 1 - \kappa(\lambda)\big], \tag{10}$$

where $\kappa(\lambda)$ is the cost of screening a patent application. We assume $\kappa(\lambda)$ is continuously differentiable, decreasing, and a convex function of the fraction of undetected obvious innovations.

The justification for (10) is as follows. In steady state, e entrants invest each period at a cost of one to produce an innovation. The innovation contributes to social welfare if it is genuine and ends up being produced, either because the entrant lands in a niche not occupied by a monopolist or because the entrant wins the patent infringement case. This event occurs with probability p. The rents associated with a successful genuine innovation are the present discounted value of a perpetual increase in quality, Π. These rents are split between firms and consumers. When a genuine innovation arrives, the innovating firm initially appropriates the rent. However, when a subsequent innovation, either genuine or obvious, is implemented in the niche, the rents of the incumbent are competed away, and the benefits of the prior increase in quality accrue to consumers.[15] In the absence of screening costs (i.e., $\kappa(\lambda) = 0$ for all λ), the free-entry condition (5) implies that the parenthesis in (10) is always positive.[16] More generally, the assumption $\kappa(1) = 0$ guarantees that, under the welfare-maximizing choice of λ, welfare remains positive.

When screening costs are zero, under exogenous courts, the derivative of the welfare function with respect to the screening rate λ is

$$\frac{\partial W}{\partial \lambda} = \frac{\partial e}{\partial \lambda} (p\Pi - 1) + e \frac{\partial p}{\partial \lambda} \Pi. \tag{11}$$

The screening level directly affects welfare through two channels: it determines the entry rate, e, and the probability of profitable entry, p. From Lemma 2, we know that p is increasing in λ. That is, for a given entry flow e, a lower screening level reduces the proportion of monopolized market niches, reducing the number of genuine innovations challenged in court and, consequently, the number of innovations that go to waste. From Lemma 3, we know that the net effect of screening λ on the entry flow e can be ambiguous

[15] Equivalently, one can compute welfare as the per-period consumer and producer surplus additions that entrants bring about, $W = \Delta CS + \Delta PS$. Ignoring screening costs, the former can be written as

$$\Delta CS = e \cdot x \cdot (\alpha\mu_1 + (1-\alpha)\lambda\mu_0)\Pi = e \cdot p \cdot \Pi,$$

since consumers receive rents from innovation when entry is successful in a monopolized niche that eliminates the rents of the incumbent. On the other hand, in steady state, entry does not alter the fraction of monopolized niches (and hence the overall surplus of producers conditional on entry). However, it implies that entrants pay an overall cost e. So $\Delta PS = -e$, implying (10).

[16] Specifically, we have $\Pi > v$ and $pv = 1$.

if $\mu_1 < 1$. This explains the next result.

Proposition 1 (Rational ignorance). *In the absence of screening costs ($\kappa(\lambda) = 0$ for all λ), there exists a threshold $\hat{\mu}_1 \in (0, 1)$ such that if $\mu_1 \geq \hat{\mu}_1$ it is socially optimal to fully screen out obvious innovations ($\lambda = 0$), and; if $\mu_1 < \hat{\mu}_1$ it is socially optimal to allow some obvious innovations. In this case, the optimal proportion of obvious innovations allowed is decreasing in Π and given by*

$$\hat{\lambda} = \begin{cases} 1 & \textit{if } \Pi < K, \\ \alpha \frac{2+(1-3\mu_1)\alpha\Pi+(1-\mu_1)\sqrt{\alpha\Pi(\alpha\Pi+8)}}{2\mu_0(1-\alpha)(\alpha\Pi-1)} & \textit{otherwise,} \end{cases} \tag{12}$$

where K is a known positive constant.

Clearly, if screening is costly and, in particular, if the screening cost function satisfies Inada-type conditions (specifically, $\kappa'(0) = -\infty$ and $\kappa(1) = \kappa'(1) = 0$), the optimal value of λ is interior. Proposition 1 goes further by saying that, even in the absence of screening costs, there may be circumstances (when $\mu_1 < 1$) in which it is optimal to grant a patent to some obvious innovations. The intuition behind this result is that obvious patents may foster entry by increasing the number of competitive niches, thus increasing the probability of success of future genuine innovators. When $\mu_1 = 1$, however, no genuine innovation goes to waste, and the benefit of increasing p is nil. Consequently, only the effect of screening on steady-state entry matters. From Lemma 3, we know that entry is decreasing in λ so that the optimal solution is full screening.

Our previous result is a version of insufficient entry à la Dixit and Stiglitz (1977), which arises when entrants cannot fully appropriate the value they generate. In our model, there are two sources of non-appropriability. First, genuine innovators get replaced by future entrants, limiting the time they can appropriate the rents they generate. Second, existing property rights act as an entry barrier, precluding some genuine innovators from appropriating any rents. This second source depends on the competitive state of the niche as, before reaching the market, innovators are uncertain whether they will face the opposition of an incumbent or not.

The screening rate by the patent office affects both mechanisms but in opposite directions. Better screening makes the entry of obvious innovators more difficult, lengthening the incumbency status and improving rent appropriability. At the same time, better screening increases the proportion of niches with an active incumbent, increasing the chances of facing opposition at the moment of entry and reducing rent appropriability. This tradeoff is uncovered in Proposition 1.

Observe that the first mechanism operates at a niche level, whereas the second mechanism operates across niches. If we ignore the barrier-of-entry effect, or if it is sufficiently small (i.e., $\mu_1 \geq \hat{\mu}_1$), the optimal screening strategy consists of thoroughly screening out obvious innovators, as explained in Green and Scotchmer (1995) and O'Donoghue et al. (1998). When we incorporate the cross-niche effects, we find that allowing some obvious innovators is optimal. To conclude this section, we solve as a benchmark the problem of a planner that can control both patents' ex-ante screening as measured by λ and their ex-post enforcement in court as represented by μ_0 and μ_1.

Corollary 1. *In the absence of screening and enforcement costs, a social planner that can decide on patents' ex-ante screening (λ) and on their enforcement (μ_0 and μ_1) chooses to fully screen out obvious innovations, either at the patent office or in court, and to always rule in favor of new genuine innovators.*

This result arises from a combination of the previous results. As shown in Lemma 2, higher values of μ_1 yield an increase in the probability that an entrant is successful, p, and, consequently, an increase in total entry, e. Both effects increase social welfare, as equation (11) indicates. Hence, a planner that could regulate the behavior of courts at no cost should choose $\mu_1 = 1$. Using Proposition 1, we know that obvious entrants should not receive any protection in that case. That is, it would be optimal to set $\lambda\mu_0 = 0$.

4 Endogenous Courts

In the previous sections, we treated as exogenous the probabilities μ_0 and μ_1 that determine the result of court trials against obvious and genuine innovators, respectively. In this section, we endogenize these probabilities as the result of an evidence-gathering process carried out by the judge who oversees each case. We first analyze the decision of an individual judge involved in a single case. We examine how this judge best responds to different ex-ante screening rates by the patent office and the expectations on the behavior of other judges that will rule on future cases. We then aggregate the decisions of all judges in the (Markov perfect) steady-state equilibrium of the model and analyze the effects of changes in the screening rate λ.

4.1 Type I and Type II Errors

We assume that when a case reaches a judge, she decides how much evidence-gathering effort s to exert in order to maximize the social welfare created in that niche.[17] The judge takes the screening rate λ and other judges' efforts as given. This means that the judge only considers the impact of that *specific* ruling on total welfare. A judge only reviews a case after entry has occurred, thus ignoring (in contrast with (10)) the already sunk cost of entry. The judge, however, takes into account the effect of the ruling on future entry. Because we focus on symmetric equilibria, we assume that all future judges will exert the same effort level $\hat{s}$.

The impact of a judge on welfare is the result of type I and type II statistical errors. We defined the cost of the type I error, E_I, as the welfare loss from precluding the production by an entrant that holds a genuine innovation. This cost is easy to assess since it reduces social welfare by π on a permanent basis, so $E_I = \Pi$, where, as defined earlier, $\Pi = \pi/(1-\beta)$.

The type II error consists of failing to protect an incumbent monopolist against an

[17] Notice that, because judges are atomistic, their decisions have no impact on the welfare outcomes of other niches.

obvious innovator, turning the monopolized niche into a competitive one. Making the niche competitive affects the probability of future entry and, consequently, the stream of future innovations. Let w_M and w_C be the present value of the welfare realized in the niche, from the following period onwards, when it is currently in the monopoly and the competitive state, respectively. The welfare losses due to the type II error can be written as $E_{II} = \beta(w_M - w_C)$, where the values w_M and w_C can be found by solving the following system of equations:

$$
\begin{aligned}
w_C &= (1 - e\alpha)\beta w_C + e\left[\alpha\left(\Pi + \beta w_M\right) - 1 - \kappa(\lambda)\right], \\
w_M &= \left[1 - e(1-\alpha)\lambda\mu_0)\right]\beta w_M + e\left[(1-\alpha)\lambda\mu_0\beta w_C + \alpha\mu_1\Pi - 1 - \kappa(\lambda)\right].
\end{aligned} \tag{13}
$$

Because judges are atomistic, they take as given the enforcement decisions of future judges, $\hat{s}$. These decisions also determine the future entry rate e and the probabilities with which these genuine and obvious innovators will prevail in court, μ_1 and μ_0, respectively.

The previous expressions can be interpreted as follows. The social value of a competitive niche before entry takes place, w_C, depends on the likelihood and quality composition of the prospective entry. With probability e, entry occurs; the entrant incurs in an entry cost of 1, and the patent office pays $\kappa(\lambda)$ for screening the entrant. A competitive niche becomes monopolized when a genuine innovator successfully enters. This event occurs with probability $e\alpha$, as the entrant does not face the opposition of an incumbent. The genuine entrant directly contributes a discounted social surplus of Π and turns the niche into a monopolized one starting the next period, generating βw_M as continuation present value of welfare. With probability $1 - e\alpha$, genuine entry does not occur, and the niche remains competitive, generating a present value of welfare βw_C.

Similarly, the social value of a monopolized niche, w_M, also depends on whether the incumbent faces entry and the entrant's identity. As before, entry takes place with probability e, generating an entry plus screening cost of $1+\kappa(\lambda)$. A monopolized niche becomes competitive, generating a continuation value of welfare βw_C, when an obvious innovator enters the niche. This event occurs with probability $e(1-\alpha)\lambda\mu_0$, as the obvious entrant needs to surmount the screening process and the incumbent's opposition. With probability $1-e(1-\alpha)\lambda\mu_0$, there is no obvious entrant, generating a continuation value of welfare βw_M. With probability $\alpha\mu_1$, the entrant is genuine and succeeds, producing an increase in social surplus Π and no change in the monopolized status of the niche.

Lemma 4 shows, by solving (13), that the type II error, E_{II} has a negative sign. Intuitively, the type II error generates a net welfare gain because the continuation social surplus increases when the niche becomes competitive: eliminating the monopoly increases the probability that a future genuine innovator enters successfully. This perceived benefit from the type II error is a reinterpretation of the classical time-inconsistency problem associated to patent policy. Whereas the ex-ante promise of protection spurs innovation and increases social welfare, once the innovation has taken place, it is optimal to prevent the exercise of the market power that a patent allows, which, in our model, does not directly produce a deadweight loss but it is detrimental to the entry of subsequent innovators and the net value they engender.

Lemma 4. *The steady-state net welfare cost associated with type II error is negative and equal to*

$$E_{II}(\hat{s},\gamma) = -\Pi\frac{(1-\gamma)\alpha\beta(1-\mu_1(\hat{s}))e(\hat{s},\gamma)}{(1-\gamma)(1-\beta)+\alpha\beta\left(1-\gamma+\gamma\mu_0\left(\hat{s}\right)\right)e\left(\hat{s},\gamma\right)} \tag{14}$$

where

$$\gamma \equiv \frac{(1-\alpha)\lambda}{\alpha+(1-\alpha)\lambda} \in [0,1-\alpha] \tag{15}$$

measures the probability that a judge faces an obvious entrant.

To simplify notation, we have implemented a change of variable from the screening quality λ to the proportion of obvious entrants faced by a judge, γ, as defined in (15). Observe that γ is increasing in λ and decreasing in the proportion of genuine innovations, α. To avoid confusion in the analysis that follows, in equation (14) we made explicit the dependency of μ_0 and μ_1 on $\hat{s}$, and of the entry rate and the cost of the type II error on $\hat{s}$ and γ. For now, we assume that judges face a given γ and take decisions based on it. The value of γ is endogeneized in Section 4.4.

4.2 An Individual Judge's Problem

We can now characterize the optimal evidence-gathering decision of a given judge, $s \in [0, 1]$. It is immediate that maximizing social welfare in a niche is equivalent to minimizing the expected social cost of both types of error plus the cost of the effort required to reduce such errors, $c(s)$. This overall cost can be expressed as

$$J(s, \hat{s}, \gamma) = (1-\gamma)(1-\mu_1(s))E_I + \gamma\mu_0(s)E_{II}(\hat{s}, \gamma) + c(s), \tag{16}$$

since, with probability $1-\gamma$ the judge faces a genuine innovation, leading to a type I error with probability $1-\mu_1(s)$, and with probability γ the judge faces an obvious innovation, leading to a type II error with probability $\mu_0(s)$.

The analysis of the decisions emerging from the minimization of (16) in the case in which s is a continuous variable and $c(s)$ is an increasing and convex cost function is rather involved. To convey intuitions in the remaining of this section, we will consider the case in which the judge's decision is binary $s \in \{0, 1\}$, and we defer the discussion of the case in which s is continuous to Section 5.2. As a result, we can normalize the cost of no effort to 0, $c(0) = 0$, and define $c(1) = c > 0$. Using (1), we have $\mu_0(0) = \mu_1(0) = \frac{1}{2}$ and $\mu_1(1) = 1 > 0 = \mu_0(1)$, which greatly simplifies the analysis.

We now turn to a judge's optimal effort decision. Under full effort, we have $J(1, \hat{s}, \gamma) = c$ for any value of γ and $\hat{s}$ since type I and type II errors do not arise and the overall social cost of the judge's decision is only the cost of her effort. The social cost when a judge exerts no effort is characterized in the next lemma.

Lemma 5. *When a judge exerts no effort, the social cost of her decision is given by*

$$J(0, \hat{s}, \gamma) = \Pi(1-\gamma)\Phi(\hat{s}, \gamma)/2. \tag{17}$$

where $\Phi(\hat{s}, \gamma)$ is a factor related to the present value of the innovation-reducing effects of type I error net of the innovation-enhancing effects of type II error, and is given by

$$\Phi(\hat{s}, \gamma) = \frac{(1-\gamma)\left[(1-\beta) + \alpha\beta e(\hat{s}, \gamma)\right]}{(1-\beta)(1-\gamma) + \beta\alpha\left((1-\gamma) + \gamma\mu_0(\hat{s})\right)e(\hat{s}, \gamma)} \in [0, 1]. \tag{18}$$

This factor is decreasing in $e(\hat{s},\gamma)$, γ, α, *and* Π.

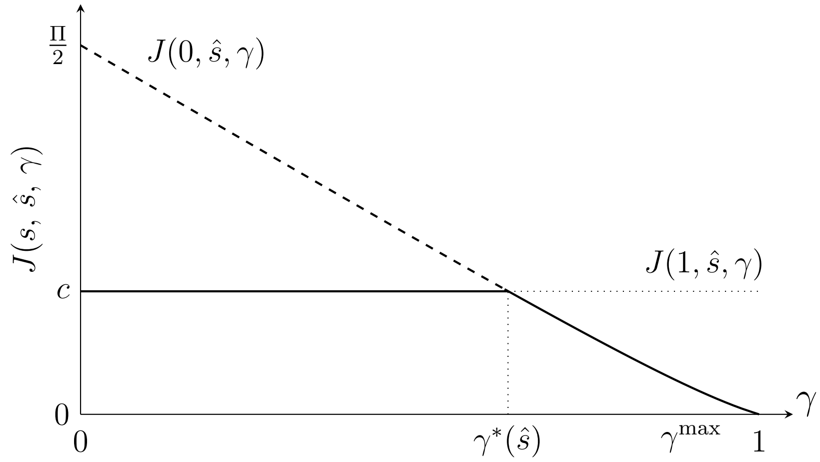
That is, for a given effort by the other judges, $\hat{s}$, and probability of facing an obvious innovator, γ, the social cost $J(0,\hat{s},\gamma)$ when a judge exerts no effort depends on the probability of facing a genuine innovator, $1-\gamma$, times the probability $1/2$ of taking the wrong decision (see (1) under $s=0$), times the magnitude of the perpetuity loss Π. The term $\Phi(\hat{s},\gamma) \in [0,1]$ measures the net intertemporal detrimental effect of type I and type II errors on the entry of genuine innovations. That is, $\Phi(\hat{s},\gamma)$ is positively affected by the type I error and negatively affected by the type II error.

Observe that $\Phi(\hat{s},\gamma) \geq 0$, meaning that the benefit of the type II error never overcomes the cost of a type I error. However, an increase in the (endogenous) entry flow raises the probability that the niche will be occupied by a genuine innovator in the future, which increases the benefits of a type II error, decreasing $\Phi(\hat{s},\gamma)$. An increase in the probability of facing an obvious entrant, γ, also increases the importance of the benefits of the type II error. Intuitively, this occurs because an increase in the proportion of obvious patents decreases the chance of future litigation, raising the probability of successful entry by a genuine innovator, which, in turn, increases future entry. Similarly, an increase in the probability of obtaining a genuine innovation α and in the discounted profits obtained by successful genuine innovators Π also boost entry, raising the importance of the benefits of a type II error and decreasing $\Phi(\hat{s},\gamma)$.

The following proposition describes how an individual judge's enforcement effort, under a given anticipated enforcement effort of other judges, depends on c and γ.

Proposition 2 (Screening complements enforcement). *For any value of other judges' enforcement effort* $\hat{s} \in [0,1]$*:* i) *If* $c < \Pi/2$*, there exists a threshold value for the probability of facing an obvious innovator,* $\gamma^*(\hat{s}) \in (0,1)$ *such that the judge exerts effort if and only if* $\gamma \leq \gamma^*(\hat{s})$. ii) *If* $c \geq \Pi/2$*, a judge does not exert effort regardless of the value of* $\gamma > 0$.

The main implication of the previous proposition is that the patent office's screening rate (which reduces λ and, hence, γ) and an individual judge's enforcement effort are



Note: Parameter values are $\alpha = 0.1$, $\beta = 0.8$, $\pi = 3.6$, $c = 3$, and $\hat{s} = 0.6$.

Figure 2: The decision of an individual judge. This figure shows the social cost $J(s, \hat{s}, \gamma)$ minimized by each individual judge as a function of the probability γ with which the confronted entrant holds an obvious innovation if all other judges choose effort $\hat{s}$. The maximum possible value of γ is $\gamma^{\max} = 1 - \alpha$.

complementary. Intuitively, suppose the patent office screens out a larger proportion of obvious innovations (lower λ). In that case, the social cost of making a judicial error increases (as it is more likely to be a type I error in this case), encouraging the judge to exert effort. Mathematically, $J(0, \hat{s}, \gamma)$ in (17) decreases in γ both directly and indirectly through $\Phi(\hat{s}, \gamma)$ as observed in Lemma 5.

Figure 2 illustrates the trade-off behind this result. It depicts the social cost internalized by an individual judge as a function of the probability of facing an obvious innovator, γ, under the two possible enforcement effort choices, $s = 0$ and $s = 1$. The social cost under full effort is flat and equal to c, whereas the social cost under no effort declines with γ. The cutoff $\gamma^*(\hat{s})$ separates the ranges of γ for which, given other judges' enforcement effort, an individual judge will or will not exert effort. The complementarity between screening rate (low γ) and enforcement effort (choice of $s = 1$) will become important when we characterize judges' behavior.

In order to focus on the case in which full enforcement effort is possible in equilibrium, henceforth, we make the following assumption:

Assumption 2. $c < \Pi/2$.

4.3 Judges' Equilibrium Enforcement Effort

The analysis in the previous section refers to an individual judge, taking the symmetric behavior of future judges, $\hat{s}$, as given. From the perspective of a judge's objective function in a niche, $\hat{s}$ only affects the type II error, and it does so through two channels. First, it affects the expected entry flow, $e(\hat{s}, \gamma)$. Second, and contingent on a given future entry flow, it affects the prospects of transition between monopoly and competition in the specific niche to which the ruling will apply. These transitions are affected by $\hat{s}$ through $\mu_0(\hat{s})$ and $\mu_1(\hat{s})$.

In a symmetric steady-state equilibrium, judges' enforcement effort s^* is given by the effort level of an individual judge s^* that is consistent with having $\hat{s} = s^*$ as the enforcement effort exerted by future judges. To characterize s^* with binary enforcement effort, it is important to understand how the threshold value of the probability of facing an obvious innovator that determines an individual judge's effort depends on $\hat{s}$.

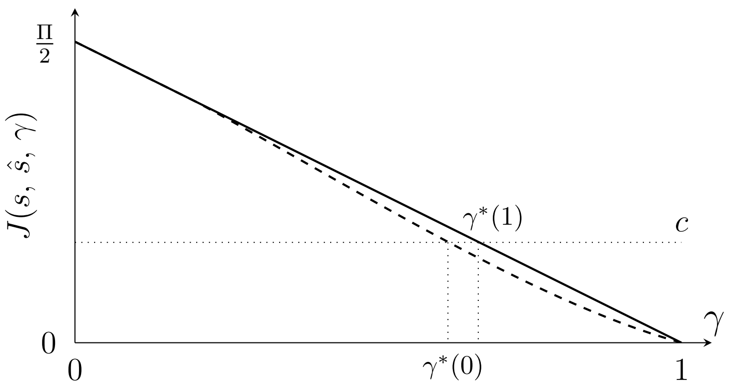
Lemma 6. *The threshold $\gamma^*(\hat{s})$ determining how a judge's effort depends on other judges' effort $\hat{s} \in \{0, 1\}$ satisfies $\gamma^*(0) < \gamma^*(1) \equiv (\Pi - 2c)/\Pi$.*

In words, the range of values of γ over which a judge exerts effort expands when future judges are also expected to exert effort, meaning that the effort of subsequent judges are strategic complements. This result, illustrated in Figure 3, yields the equilibrium configurations described in the next proposition.

Proposition 3 (Enforcement equilibria)**.** *In a pure-strategy symmetric steady-state equilibrium, judges' effort in the patent enforcement game is given by:*

$$s^* = \begin{cases} 1 & \textit{if } \gamma \leq \gamma^*(0), \\ \{0, 1\} & \textit{if } \gamma \in (\gamma^*(0), \gamma^*(1)], \\ 0 & \textit{if } \gamma > \gamma^*(1). \end{cases}$$

In the multiple equilibrium region, entry is higher with full enforcement effort ($s^* = 1$) than with no enforcement effort ($s^* = 0$).



Note: Parameter values are $\alpha = 0.1$, $\beta = 0.8$, $\pi = 3.6$, and $c = 3$.

Figure 3: Judges' equilibrium enforcement effort. The solid and dashed line represent $J(0, 1, \gamma)$ and $J(0, 0, \gamma)$, respectively. An equilibrium with $s^* = 1$ ($s^* = 0$) exists whenever the probability of facing an obvious innovator is below (above) $\gamma^*(1)$ ($\gamma^*(0)$)

A judge exerts effort only if the patent office's screening is sufficiently thorough. However, the specific threshold where this complementarity affects a judge's behavior depends, in turn, on the decisions of future judges. To the extent that a choice $\hat{s} = 0$ ($\hat{s} = 1$) by other judges strengthens the incentives for a given judge also to choose $s = 0$ ($s = 1$) a second source of complementarity arises. As is often the case, this effect leads to a coordination game and multiplicity of equilibria. With low effort by future judges, $\hat{s} = 0$, some obvious innovators end up replacing existing monopolists, which contributes to increasing the social value of not exerting effort and facilitating the entry of future genuine innovators. However, when future judges exert effort, $\hat{s} = 1$, this entry-facilitating effect of $s = 0$ (and, thus, the convenience of the type II error) vanishes because future judges always allow genuine innovators to enter. Without a type II error, whether the judge chooses $s = 0$ or $s = 1$ only depends on comparing the gains from reducing type I error with the enforcement effort cost c.

We can now characterize how the equilibrium enforcement effort changes with the parameters of the model.

Proposition 4 (Comparative Statics)**.** *In the symmetric steady-state equilibrium, judges' enforcement effort s^* is increasing in the value of a genuine innovation Π, increasing in the* *screening quality of the patent office (decreasing in λ), and decreasing in the enforcement effort cost, c.*

In the previous section, we established the complementarity between the effort carried out by an individual judge and the ex-ante screening of patents. This result naturally extends to the equilibrium enforcement effort level. A lower λ, arising from a more thorough application review process by the patent office and fewer patents granted for obvious innovations, implies a lower value of γ and a higher enforcement effort by all judges.

This result also allows us to understand under which circumstances the level of patent enforcement that maximizes entry and innovation (characterized in Proposition 1 for cases where courts were assumed to be exogenous) can be attained once courts are endogenized. As it turns out, the complementarity between ex-ante screening and ex-post judge effort implies that λ has to be sufficiently small so that $\gamma \leq \gamma^*(1)$ in order to induce $\mu_1(1) = 1$ and $\mu_0(1) = 0$. In the next section, we build on this result and explore the socially optimal choice of λ once its impact on judges' enforcement efforts is taken into account.

The previous proposition also establishes a positive relationship between the value of an innovation and judges' equilibrium enforcement effort. For each individual judge, and given $\hat{s}$, an increase in π or β generates two effects that operate in opposite directions. On the one hand, and as established in Lemma 5, if the value of the (future) innovation increases, allowing obvious innovations today decreases the cost of $s = 0$ due to the higher social value of type II error. On the other hand, the higher the discounted value of the innovation, the higher the cost of mistakenly preventing the production of a genuine innovation in the current period (type I error). When π or β increase, this second effect dominates, implying an upward shift in the function $J(0, s^*, \gamma)$ for both $\hat{s} = 0$ and $\hat{s} = 1$. In terms of Figure 3 this means that both $\gamma^*(0)$ and $\gamma^*(1)$ increase, expanding the region over which an equilibrium with $s^* = 1$ is sustainable. As genuine innovations become more prevalent (lower λ) and their social value increases (higher $\Pi = \frac{\pi}{1-\beta}$), investing in enhancing the quality of enforcement becomes more valuable, and an equilibrium with full enforcement is more likely to emerge.[18]

4.4 Screening and Enforcement Equilibrium

In this section, we characterize the socially optimal patent screening rate, λ^*, taking into account the endogenous response of courts. Given the equilibrium enforcement effort decision of judges characterized in the previous section, $s^*(\lambda, c)$, a social planner chooses λ to maximize[19]

$$W(\lambda; s^*, c) = e(s^*, \lambda)\left[p(s^*, \lambda)\,\Pi - 1 - \kappa(\lambda) - c \cdot s^* \cdot (\alpha + (1-\alpha)\lambda) \cdot x(s^*, \lambda)\right], \quad (19)$$

where $x(s^*, \lambda)$, $p(s^*, \lambda)$, and $e(s^*, \lambda)$ are the equilibrium values for the proportion of monopolized niches, the probability that an entrant obtains a genuine innovation, and the number of entrants, respectively (see equations (6), (7) and, (8)). Social welfare differs from the patent's office problem (10) in two ways. First, in equation (10), judges' anticipated enforcement effort was exogenously fixed, while in (19), the social planner accounts for the effect of λ on s^*. Second, social welfare in (19) also takes into account that judges incur in a cost of $c \cdot s^*(\lambda, c)$ per case reviewed and that the mass of cases reviewed is the proportion $(\alpha + (1-\alpha)\lambda)\, x(s^*, \lambda)$ of entrants that obtain a patent and reach a monopolized niche.

To solve the planner's problem, it is convenient to define the socially-optimal screening values as a function of the enforcement cost c, conditional on each possible level of enforcement effort. Let $\lambda_0 \equiv \arg\max_{\lambda \in [0,1]} W(\lambda; s = 0)$, which does not depend on c, since when judges exert no effort, enforcement costs do not affect social welfare. Similarly, let $\lambda_1(c) \equiv \arg\max_{\lambda \in [0,1]} W(\lambda; s = 1, c)$. The proof of Proposition 5 shows that these objects are well-defined.

[18]The effect of α is more difficult to assess. It is true that, given γ, the cost of poor enforcement decreases as α increases (see Lemma 5). However, the value of γ negatively depends on α. Numerical calculations indicate that, as in the case of Π, the overall effect of α on s^* is positive, although an analytical characterization of this result is elusive.

[19]To simplify the notation, we omit λ and c from $s^* = s^*(\lambda, c)$ as arguments in equation (19).

Assumption 3. *The screening cost $\kappa(\lambda)$ satisfy Inada conditions: $\kappa'(\lambda) \leq 0$, $\kappa''(\lambda) > 0$, $\kappa'(1) = 0$, and $\lim_{\lambda \to 0} \kappa'(\lambda) = -\infty$. Also, Π and $\kappa(\lambda)$ satisfy $\frac{2[(1-\alpha)c]^2}{\alpha \kappa''(\lambda_1(c))} < \Pi$ for $c < \Pi/2$.*

This assumption is well defined as $\lambda_1(c)$ does not depend on Π. The assumption guarantees that $\lambda_1'(c)$ is not too steep, leading to a well-behaved solution.[20]

Proposition 5 (Optimal Screening). *There exists an enforcement cost threshold c^* inducing full judiciary effort when $c \leq c^*$ and no effort otherwise. The socially optimal patent screening rate, $\lambda^*(c)$, decreases in $c < c^*$, has a upward discontinuity at c^* (i.e., $\lambda^*(c^*-) \leq \lambda^*(c^*+)$), then decreases again for $c > c^*$ until it reaches λ_0, becoming constant.*

Proposition 5 shows that patent screening and enforcement can be strategic complements or substitutes from a social welfare perspective. The complementarity is captured at the threshold c^*. There, increasing enforcement costs induce judges to stop exerting effort, which is accommodated by a decrease in the optimal screening quality—the benefits of patent screening are lower when no judiciary effort is present (see Figure 4). The substitutability between screening and enforcement is captured when enforcement costs are low enough to induce positive judicial effort. In that scenario, increasing enforcement costs leads to better patent screening by the patent office (a lower $\lambda^*(c)$); better screening reduces the number of cases judges review, compensating for the increase in enforcement costs. Section 5.2 studies a model with continuous enforcement efforts in which the strategic complementary and substitutability forces are always at play. We show that in that case, the complementarity is the dominant force.

Finally, with high enforcement costs, it is socially optimal for the judges to exert no effort. In this case, screening costs do not directly enter the welfare function (19), and the optimal screening rate is λ_0. This ideal screening, however, might not be feasible as it may induce judges to exert effort. The optimal social screening is, thus, constrained by the equilibrium behavior of judges to the highest screening level inducing no effort. This constrained screening decreases in c until it reaches λ_0. At that point, λ_0 becomes

[20]For example, the cost function $\kappa(\lambda) = (1-\lambda)^2/\lambda$ satisfies this condition for sufficiently large Π.

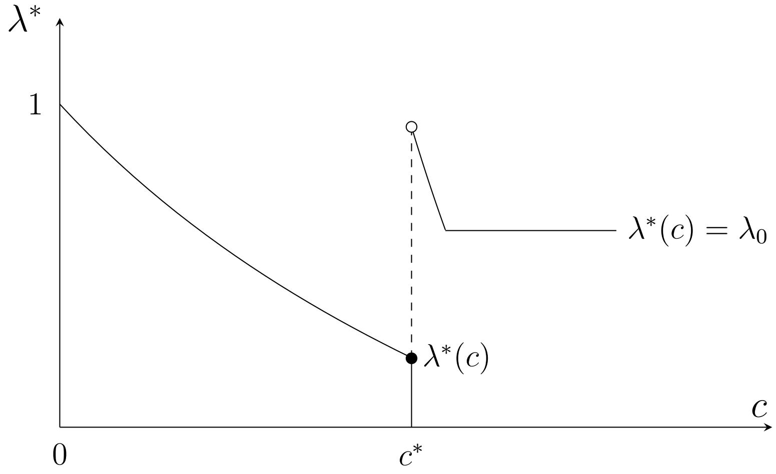


Figure 4: Optimal screening as a function of the enforcement cost. Screening increases ($\lambda^*(c)$ declines) with the enforcement cost c while judges exert high enforcement effort. When $c > c^*$, judges perform no effort and, due to the complementarity between screening and enforcement, screening jumps down.

feasible, and the socially optimal screening rate stays constant with increased enforcement costs.

5 Robustness and Extensions

Here we develop several extensions of the baseline model and show that the main insights obtained in previous sections are robust to these changes. First, we consider the case in which innovation reduces production costs and demand is not unit-inelastic, in which case the model features static inefficiency. Second, we discuss the case in which enforcement effort is continuous. In the third extension we show the robustness of the result to generalizing the probability with which judges rule against entrants when exerting no enforcement effort. The fourth extension develops the case in which the knowledge associated with blocked genuine innovations is not lost and future genuine innovators can build on it. The fifth and final extension discusses the implications of allowing the incumbents of competitive niches (i.e. successful obvious innovators) to extract rents from subsequent innovators by acting as patent trolls. Technical details are relegated to Appendix B.

5.1 Cost-saving Innovations and Static Inefficiency

The baseline model considers quality-improving innovations in an environment with inelastic demands. In that context, genuine innovators extract all the surplus from consumers, avoiding deadweight losses and simplifying the welfare analysis. We now provide a tractable framework that extends our analysis to a case of cost-reducing innovations where market power involves welfare losses.

In this setup, there is also a continuum of niches of size 1. In each niche, the good produced is homogeneous. Firms compete in prices and face a demand function $q = a/p$. Each genuine innovation decreases the existing marginal cost by a proportion $1-\delta$ where $\delta \in (0,1)$. That is, if z represents the baseline marginal cost, after m genuine innovations the marginal cost becomes $z_m = \delta^m z$.[21]

Lemma 7. *In a cost-saving innovations setup, the profit flow π and the deadweight loss ℓ generated by a genuine innovation are independent of the baseline marginal cost z and the cumulative number of innovations m. In particular, as illustrated in Figure 5, they are equal to $\pi = a(1-\delta)$ and $\ell = a\left(\ln(\delta^{-1}) - (1-\delta)\right) > 0$.*

Social Welfare Because profits are invariant to the cumulative number of innovations, firm behavior and industry dynamics described in Section 3 go through without alterations. The objective functions of the patent office and courts, however, need to be reformulated to account for the deadweight loss ℓ. In particular, the main difference with respect to the baseline model is that now, allowing entry into a monopolized niche converts the existing deadweight loss into consumer surplus. Entry, whether from a genuine or obvious entrant, increases the social welfare by ℓ on a permanent basis, $L \equiv \ell/(1-\beta)$. As carefully shown in Appendix B.1, the objective function of the patent office is now

$$W = e\left[p(\Pi + L) - 1 - \kappa(\lambda)\right].$$

[21]This demand and the proportional cost-saving innovation are also used in Marshall and Parra (2019).

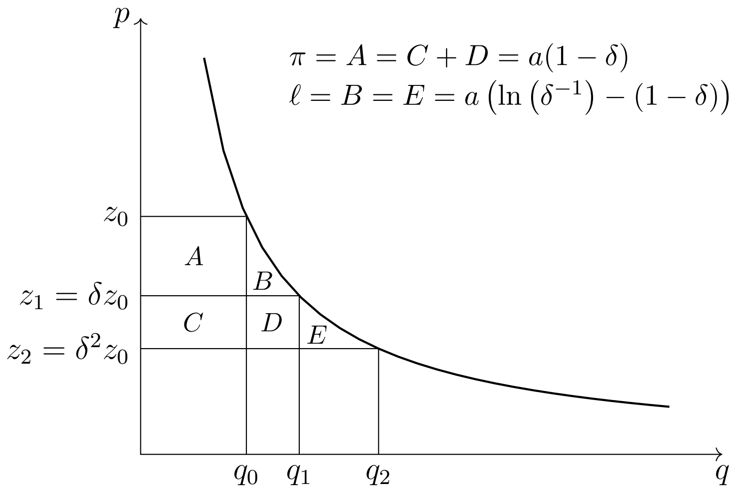


Figure 5: Monopoly profits and deadweight losses with cost-saving innovations. This figure identifies the profits (π) and deadweight losses (ℓ) associated with a successful genuine cost-saving innovation as areas under the relevant demand curve.

This expression is analogous to that in (10) except for the new term in L, which captures the deadweight loss recouped with the arrival of an innovation. Because the behavior of e and p with respect to λ remains unchanged with respect to the baseline model, it is immediate that the message of Proposition 1 extends to this environment. When courts' behavior, as represented by μ_0 and μ_1, is exogenous, the optimal patent screening rate may be interior even in the absence of screening costs.

The Judges' Problem We can now analyze how a judge's enforcement decision changes when innovations are cost reducing and monopoly involves a deadweight loss. Recall that entry is only opposed by the incumbent in a monopolized niche. Therefore, a ruling in favor of the entrant will always increase welfare by (at least) L regardless of the entrant's innovation quality. As in the benchmark case, a type I error arises whenever a genuine innovation is excluded from the market. Since this error can only occur in already monopolized niches, it now leads to a loss of $E_I^{CS} = \Pi + L$, where the superindex CS stands for the cost-saving innovation setup.

The cost of type II error represents the reduction (in fact, a gain, since its sign is negative as in the benchmark model) in social welfare that occurs when a firm with an obvious innovation is allowed to replace an active monopolist. This error now has two components. First, each time a monopolist is replaced by another firm, the deadweight loss associated to its innovation is eliminated, leading to an immediate gain of L. Second, as in the benchmark case, the dynamic effect of enhanced entry increases the future social value of the niche. Altogether, the benefit of type II error is now larger compared to the benchmark model; that is, for every $\hat{s}$ and γ, $E_{II}^{CS}(\hat{s},\gamma) < E_{II}(\hat{s},\gamma)$.

As before, an individual judge decides her enforcement effort s taking as given the enforcement effort decisions of other judges $\hat{s}$. We denote as $J^{CS}(s,\hat{s},\gamma)$ the cost minimized by the individual judge in this case. As in the benchmark case, when a judge chooses $s=1$, no error is made, so the only cost is that related to her effort, $J^{CS}(1,\hat{s},\gamma)=c$. When a judge chooses effort $s=0$, however, the cost is

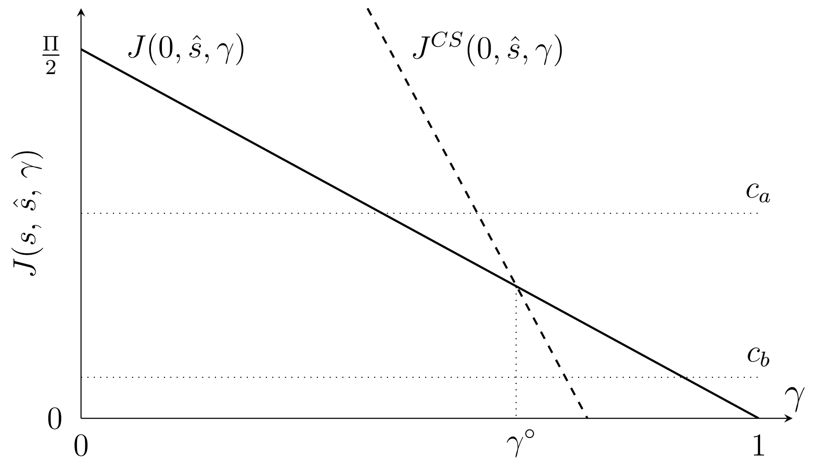
$$J^{CS}(0,\hat{s},\gamma) = J(0,\hat{s},\gamma) + L(1-\gamma)\Delta(\hat{s},\gamma)/2,$$

where $J(0,\hat{s},\gamma)$, defined in (17), is the judge's loss function in the model without deadweight losses and

$$\Delta(\hat{s},\gamma) = \frac{(1-2\gamma)(1-\beta) + \alpha\beta(1-\gamma)e(\hat{s},\gamma)}{(1-\gamma)(1-\beta) + \alpha\beta\left(1-\gamma+\gamma\mu_0(\hat{s})\right)e(\hat{s},\gamma)} \leq 1$$

is a function which, given the effort of other judges and the level of screening by the patent office, measures the net intertemporal contribution of type I and type II errors to the occurrence of the deadweight losses represented by L.

Similarly to the factor $\Phi(\hat{s},\gamma)$ that measures the net detrimental impact of judicial errors on the entry of genuine innovators, the factor $\Delta(\hat{s},\gamma)$ is decreasing in the probability that the confronted entrant is an obvious innovator, γ, which implies that the presence of the deadweight loss L reinforces the complementarity between patent screening (which reduces γ) and patent enforcement (which avoids the cost $J^{CS}(0,\hat{s},\gamma)$). Observe that, for values of γ close to one $\Delta(\hat{s},\gamma)$ can be negative, whereas when γ is low $\Delta(\hat{s},\gamma)$ is positive. The stronger dependence of $J^{CS}(0,\hat{s},\gamma)$ on γ explains the following result:



Note: Parameter values are $\alpha = 0.1$, $\beta = 0.8$, $a = 4$, $\delta = 0.1$, $c_a = 5$, $c_b = 1$ and $\hat{s} = 1$.

Figure 6: Costs minimized by an individual judge with and without static inefficiency. The solid and dashed lines represent $J(0, \hat{s}, \gamma)$ and $J^{CS}(0, \hat{s}, \gamma)$, respectively. The parallel dotted lines show two possible illustrative levels of the cost of high enforcement effort.

Proposition 6 (Enforcement incentives with deadweight losses)**.** *For every $\hat{s} \in [0, 1]$, there exists $\gamma^{\circ}(\hat{s}) \in (0, 1)$, increasing in $\alpha\Pi$, such that: i) when $\gamma > \gamma^{\circ}(\hat{s})$ a judge has less incentives to exert effort relative to a situation without static inefficiency; ii) when $\gamma < \gamma^{\circ}(\hat{s})$ a judge has more incentives to exert effort relative to a situation without static inefficiency.*

Thus, as shown in Figure 6, whether a judge exerts more or less effort relative to the model without deadweight losses depends on the values of γ and c. As in the baseline case, provided c is not too large, the judge exerts effort if and only if γ is low enough. For high values of the enforcement cost, such as c_a, the range of values of γ leading to maximum enforcement effort is wider in the situation with deadweight losses. But when the cost is low enough, such as c_b, the result is reversed, and, in the presence of deadweight losses, maximum effort prevails over a narrower set of values of γ.

From here, the final characterization of the possible outcomes of judges' enforcement game would be analogous to that in Proposition 3. Relative to the model without deadweight losses, the ranges of values of γ over which an equilibrium with high enforcement effort can be sustained expand with c.

5.2 Continuous Enforcement Effort

From Section 4.2 onwards, we streamlined the analysis by focusing on the case in which judge enforcement effort can take only two values, $s \in \{0, 1\}$. In this section, we explore the case where effort is continuous, $s \in [0, 1]$. As expected, the main results carry through. To ease the exposition, we assume that the cost of exerting effort is quadratic. In particular $c(s) = \tilde{c}s^2/2$ where $\tilde{c} > 0$ is a scale parameter. In line with Assumption 2 above, in this section, we focus on the case in which effort is interior, that is, $\tilde{c} > \Pi/2$. We provide details for this case and for the case where $\tilde{c} \leq \Pi/2$ in Appendix B.2.

For a given enforcement effort by other judges $\hat{s}$ and screening rate by the patent office, γ, a judge's best response is unique and given by

$$s(\hat{s}, \gamma) = \frac{\Pi(1-\gamma)}{2\tilde{c}} \Phi(\hat{s}, \gamma) \in (0, 1)$$

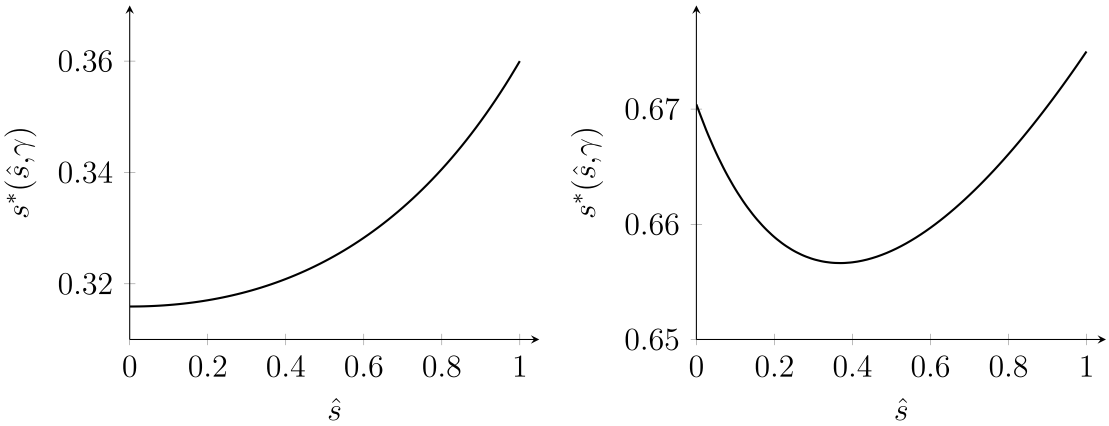
where $\Phi(\hat{s}, \gamma)$ is the function defined in Lemma 5. It also follows from this lemma that the best response of an individual judge, $s(\hat{s}, \gamma)$, increases in the quality of the screening of the patent office (decreases in γ). That is, ex-ante patent screening remains complementary to ex-post enforcement under continuous enforcement effort.

Proposition 7 (Revisiting judges' effort complementarity)**.** *In the continuous enforcement effort case, a judge's best response is increasing in the enforcement effort of other judges $\hat{s}$ if and only if*

$$\Pi > \frac{3 - \hat{s} - 2\gamma}{\alpha(1 + \hat{s}(1 - 2\gamma))} \tag{20}$$

This condition always holds in a neighborhood of $\hat{s} = 1$ or $\gamma = 1$.

Proposition 7 shows that, unlike in the binary effort scenario, where judge enforcement effort decisions are complementary, here judges' enforcement efforts can be strategic complements or strategic substitutes. An increase in other judges' effort $\hat{s}$ induces two opposing effects. As in the baseline model (see Lemma 2), the proportion of monopolized



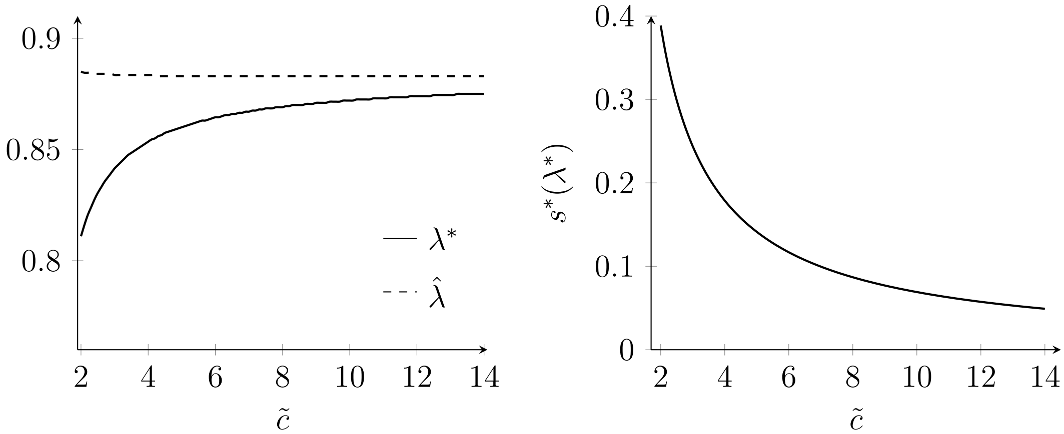
(a) Strategic complements ($\gamma = 0.6$) **(b)** Substitutes and complements ($\gamma = 0.25$)

Note: Parameter values are $\alpha = 0.1$, $\beta = 0.8$, $\pi = 3.6$, and $\tilde{c} = 14$.

Figure 7: Complementarity in the continuous enforcement effort case. Panels show the best response of an individual judge to other judges' enforcement effort $\hat{s}$ for two different levels of the screening by the patent office, as (inversely) measured by γ. In panel (a) judges' efforts are strategic complements, while in panel (b) they can be strategic substitutes or complements depending on the size of $\hat{s}$.

niches increases, increasing the frequency of type I errors relative to type II errors. This effect increases the factor $\Phi(\hat{s}, \gamma)$, calling for more effort, that is, pushing for strategic complementarity. On the one hand, in the continuous effort scenario, a new free-riding effect arises. Better (but not perfect) rulings by other judges increase the probability of future successful entry, $p(\hat{s}, \gamma)$. The prospect of higher entry increases the benefits of a type II error, decreasing $\Phi(\hat{s}, \gamma)$ (see Lemma 5) and, through it, the effort of an individual judge's best response. This effect (which vanishes in the proximity of $\hat{s} = 1$) pushes towards strategic substitutability.

The necessary and sufficient condition (20) tells us that strategic complementarity tends to occur when incumbency profits are sufficiently high. The condition becomes weaker when the patent office screens less (higher γ) and when other judges exert more effort (they are already making good rulings). Consistent with the binary effort case, regardless of the model's parameters, strategic complementarity always occurs when other judges' enforcement effort is high enough. Complementarity also occurs when the patent office screens little and/or most innovations are obvious (that is, when γ is high); see



(a) Optimal screening

(b) Optimal enforcement

Note: parameter values are $\alpha = 0.1$, $\beta = 0.8$, and $\pi = 3.6$.

Figure 8: Optimal screening and induced level of enforcement in continuous effort case. Panel (a) depicts optimal ex-ante screening when enforcement effort is endogenous and exogenous (λ^* and $\hat{\lambda}$, respectively), as a function of the enforcement cost $\tilde{c}$. Panel (b) shows the enforcement effort $s^*(\lambda^*)$ induced by the optimal ex-ante screening quality λ^*, also as a function of $\tilde{c}$.

Figure 7a. In contrast, as depicted in Figure 7b, substitutability can emerge under higher ex-ante screening.

In the Appendix, we show that the multiplicity of equilibria in the enforcement game is not present when judges' enforcement effort is continuous. That is, there is a unique symmetric steady-state equilibrium with an effort level s^* satisfying $s(s^*, \gamma) = s^*$. This means that, with continuous enforcement efforts, the judges' enforcement game no longer entails a coordination problem.[22]

To conclude this section, we numerically explore the socially-optimal screening (see problem (19)) under continuous enforcement effort. Figure 8 shows a case in which, consistent with the complementarity discussed in Proposition 5, an increase in the enforcement costs decreases both judges' enforcement effort (panel (b)) and patent office's ex-ante screening (panel (a)). That is, despite the substitution effect calling for better

[22]This suggests that the coordination problem in the binary efforts case is related to the manner in which the prospects of perfect enforcement by subsequent judges ($\hat{s} = 1$) fully removes the social value of type II error, reinforcing the incentive of an individual judge to choose $s = 1$. For $\hat{s} < 1$, type II error by an individual judge is still valuable at the margin.

screening when the cost of enforcement goes up, the impact of a decrease in enforcement effort dominates, inducing less screening in equilibrium.

To illustrate this complementarity in a different way, we also compare the socially optimal ex-ante screening behavior λ^*, just described, with what would be the optimal screening in the hypothetical case in which court behavior remains exogenously fixed at the equilibrium value induced by λ^*, $s^*(\lambda^*)$. We call the optimal screening rate in the hypothetical scenario $\hat{\lambda}$. As shown in Figure 8a, when courts' endogenous responses are taken into account, it is socially efficient to screen more (to set a lower λ): the possibility to affect court's behavior, induces more screening in equilibrium.

5.3 Different Judge Priors

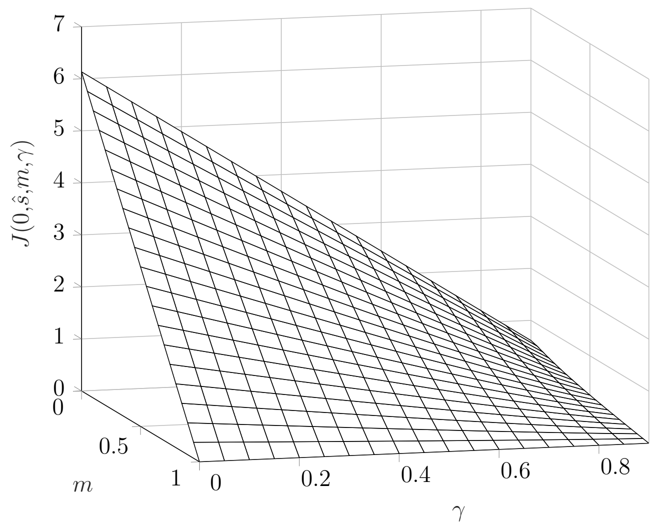
In the benchmark model, the functional form assumed in (1) implies that under no effort, $s = 0$, a judge would rule with the same probability in favor of an entrant or the incumbent, regardless of the kind of innovation that the former possessed, $\mu_0(0) = \mu_1(0) = \frac{1}{2}$. We adopted that functional form for analytical convenience, but, as shown below, its specific properties at $s = 0$ have no meaningful bearing on our results.

Consider now a more general function for the probability that a judge rules in favor of the entrant,

$$\mu_0(s, m) = m(1 - s) \quad \text{and} \quad \mu_1(s, m) = m + (1 - m)s, \tag{21}$$

where $m \in (0, 1)$ is a new parameter capturing how pro-entrant are the priors of the judge (that is, the propensity to rule in its favor under no screening). In the benchmark model, $m = \frac{1}{2}$.

In this more general case, the type I error is unchanged. Similarly, the results in Proposition 4 that determine the size of the type II error still apply, as the expressions were characterized for generic functions μ_0 and μ_1. Looking at a judge's cost of making the incorrect decision (16), we can see that raising m increases the weight given to type II errors. This effect decreases the judge's cost of not exerting effort, but the composition



Note: Parameter values are $\alpha = 0.1$, $\beta = 0.85$, $\pi = 2.3$, and $\hat{s} = 0.6$.

Figure 9: Judge's cost of no effort under different pro-entrant prior m. This figure shows the social cost $J(s, \hat{s}, \gamma)$ is decreasing in the judge's prior m and its probability of confronting an entrant holding an obvious innovation γ.

of the type II error also changes with m. The next lemma shows that, for any value $\hat{s}$ and m, the social cost for the judge of exerting effort continues to decrease in γ.

Lemma 8. *For any m, the social cost of a judge exerting effort, $s = 0$, is given by*

$$J(0, \hat{s}, \gamma) = \Pi(1-\gamma)(1-m)\Phi(\hat{s}, m, \gamma),$$

where $\Phi(\hat{s}, m, \gamma) > 0$ is given by (18), which is decreasing in γ.

This result is a counterpart of Lemma 5 and, using the same reasoning as in the benchmark model, it guarantees that in the binary effort case, $s \in \{0, 1\}$, the equilibrium choice of the judge is well defined, and induces the same behavioral implications: the judge's effort is decreasing in the probability of facing an obvious entrant, γ. Consistent with the intuition above, Figure 9 numerically shows that the social cost of not exerting effort decreases when the judge holds a more pro-entrant prior.[23] In equilibrium, a more pro-entrant judge responds by reducing the likelihood of exerting effort, relying more on its prior.

[23]This result can also be shown analytically, at least for $m \geq \frac{1}{2}$. The proof is highly involved and available from the authors upon request.

5.4 The Knowledge from Blocked Innovations Persists

In the benchmark model, genuine innovations blocked by judges are wasted, making the type-I error's cost equal to the innovation's present value. We now consider a scenario in which future entrants can build upon the knowledge of the blocked patent. In this situation, the incumbent can still enjoy the profits from its original monopoly, but it does so under the risk of entry of a firm with a genuine innovation that would now be two quality steps ahead.

5.4.1 Profits and Law of Motion of Niches

We need to extend the baseline model to accommodate an *interim* stage for incumbents that successfully defended themselves from genuine entrants. Let us denote the incumbent's state by the superscripts $\{M, I\}$ representing the monopoly and interim states, respectively. The incumbent's value in the monopoly state can be written as

$$v_t^M = \pi + \beta \left\{1 - e_{t+1}\left[(1-\alpha)\lambda\mu_0 + \alpha\right]\right\} v_{t+1}^M + \beta e_{t+1}\alpha(1-\mu_1)v_{t+1}^I.$$

Differently from equation (2), now when an incumbent successfully defends against a genuine entrant, it moves to the interim state, obtaining a value

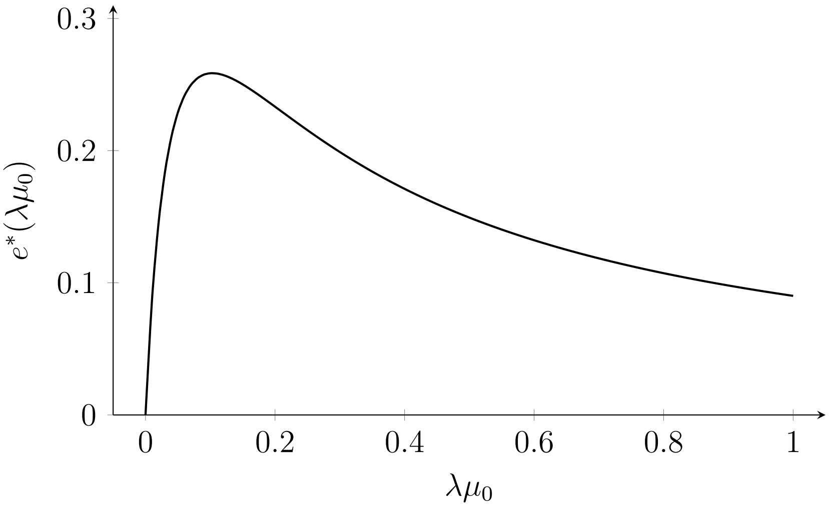
$$v_t^I = \pi + \beta \left\{1 - e_{t+1}\left[(1-\alpha)\lambda\mu_0 + \alpha\right]\right\} v_{t+1}^I.$$

In the interim state, the incumbent receives profits until successful entry occurs. As in the baseline model, we assume that the incumbent may fend off obvious innovators in court. However, when faced with a genuine innovator, the incumbent loses, as judges can appreciate the large gap between the technologies.

This extended model now has two state variables, x_t^M and x_t^I, representing the proportion of niches in the monopoly and the interim states, respectively. Their laws of motion are

$$x_{t+1}^M = x_t^M\left[1 - e_{t+1}((1-\alpha)\lambda\mu_0 + \alpha(1-\mu_1))\right] + (1 - x_t^M)\alpha e_{t+1},$$

$$x_{t+1}^I = x_t^I\left\{1 - e_{t+1}\left[(1-\alpha)\lambda\mu_0 + \alpha\right]\right\} + x_t^M\alpha(1-\mu)e_{t+1}.$$



Note: Parameter values are $\alpha = 0.1$, $\beta = 0.8$, $\pi = 2.8$, and $\mu_1 = 0.6$.

Figure 10: Entry flow and the protection against obvious innovators. This figure shows a case in which entry is maximized at an interior value of the probability with which obvious innovators are allowed to enter, $\lambda\mu_0$.

The proportion of monopolized niches increases when a genuine innovator arrives at a niche in the competitive or interim state and decreases when obvious innovators succeed against monopolist incumbents. Interim niches become competitive when obvious innovators succeed in court and become monopolized with the arrival of genuine innovators. The proportion of interim niches increases when incumbents in monopolized niches succeed in court against genuine innovators. Finally, reflecting the effective lack of opposition to genuine entry in interim niches, the probability of successful genuine entry takes the same form as in the benchmark model where the relevant stock of incumbents is x_t^M. That is, $p_t = \alpha(1 - x_t^M + x_t^M \mu_1)$ as in (4).

This version of the model embeds the same trade-off regarding the effects of patent policy as the baseline setup. The results under exogenous enforcement remain qualitatively unchanged, as can be observed, for example, in Figure 10, which constitutes a counterpart to Figure 1 and shows that entry has an inverse U-shape with respect to the level of screening against obvious innovations.

5.4.2 Endogeneous Courts

The Bellman equations describing social value in the steady state must now account for the value associated with each of the three types of niches: monopolized, in the interim state, and competitive.

$$
\begin{aligned}
w_C &= (1-e\alpha)\beta w_C + e\left[\alpha(\Pi+\beta w_M) - 1 - \kappa(\lambda)\right], \\
w_I &= \left[1 - e\left(\hat{\lambda}+\alpha\right)\right]\beta w_I + e\left[\alpha(2\Pi+\beta w_M) + \hat{\lambda}(\Pi+\beta w_C) - 1 - \kappa(\lambda)\right], \\
w_M &= \left[1 - e\left(\hat{\lambda}+\alpha(1-\mu_1)\right)\right]\beta w_M + e\left[\hat{\lambda}\beta w_C + \alpha\mu_1\Pi + \alpha(1-\mu_1)\beta w_I - 1 - \kappa(\lambda)\right],
\end{aligned}
$$

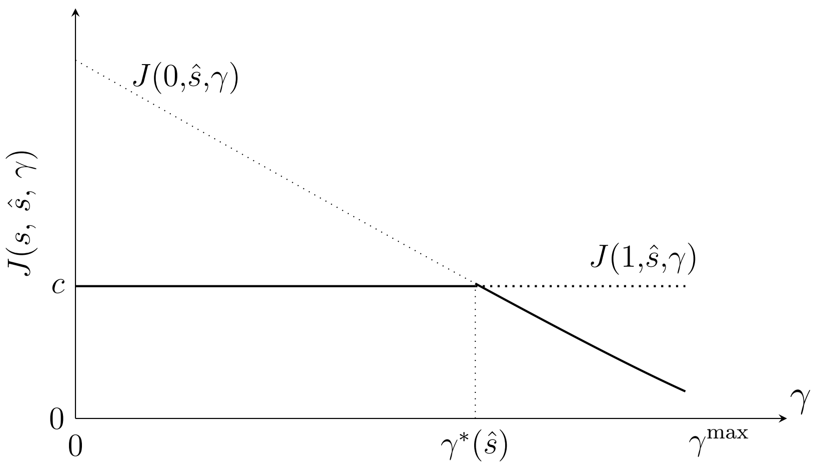
where $\hat{\lambda} \equiv (1-\alpha)\lambda\mu_0$. The expressions here are parallel to those in (13). They take into account that genuine entry into interim niches brings the social gains from both the previously blocked and the new genuine innovations (and the transition to state M). Successful obvious entry, in turn, leads to the realization of the gains from the previously blocked innovation (and the transition to state C).

In this setup, we now show that our baseline results remain qualitatively unchanged. The main difference is the cost of the type I error, which now becomes $E_I = \Pi + \beta(w_M - w_I)$, since blocking a genuine innovation now implies moving the niche to the interim stage, forgoing the continuation value w_M but gaining the value w_I. In the interim state, the blocked innovation value is delayed until entry occurs in the niche and the protection against further genuine innovation disappears, so the niche becomes more likely to accommodate genuine entry. As in the benchmark model, the cost of the type II error is $E_{II} = \beta(w_M - w_C)$.

The following results show that the basic properties of the baseline model still apply.

Proposition 8. *In the steady state of the model with an interim state, if $e > 0$ and $\beta < 1$, then $\Pi > E_I > 0 \geq E_{II}$, with $E_{II} < 0$ if $\mu_1 < 1$. If enforcement effort is binary $s \in \{0,1\}$, a judge's cost is $J(0,\hat{s},\gamma) > 0$ for all $\hat{s}$ and γ, and $J(1,\hat{s},\gamma) = c$, as in the baseline model.*

Thus, even when blocked innovations are only delayed when a judge does not uphold a genuine patent, the complementarity of screening and enforcement still goes through.



Note: Parameter values are $\alpha = 0.1$, $\beta = 0.80$, $\pi = 2.8$, $c = 2$ and $\hat{s} = 0.6$.

Figure 11: An individual judge's decision when blocked innovations get delayed. This figure shows the social cost $J(s, \hat{s}, \gamma)$ minimized by each individual judge as function of the probability γ with which the confronted entrant holds an obvious innovation if all other judges choose effort $\hat{s}$. The maximum possible value of γ is $\gamma^{\max} = 1 - \alpha$.

Figure 11 numerically shows that a sufficiently large screening of obvious patents (low γ) induces judges to exert effort.

5.5 Patent Trolling

In the benchmark model, niches occupied by incumbents with obvious patents were competitive and assumed to create no barrier to entry. We justified this assumption on the absence of incentives for producers in a competitive niche to legally defend a market position that generates no rents. However, this assumption implies abstracting from patent trolling: the use of undeserved patents as a means to extract rents from genuine innovators.

In this section, we explore the opportunistic behavior that arises when these patent trolls induce an entrant to pay a licensing fee to avoid being legally challenged in a world with imperfect courts where, otherwise, a genuine innovator faces the risk of being mistakenly blocked. This behavior, which we denote as *trolling*, is a transfer from genuine entrants to trolls, which reduces the incentives for genuine innovators to enter. Beyond that, the equations, mechanisms determining welfare, and the costs of judges' errors remain unchanged.

In the model with trolling, a genuine incumbent receives the same value, v_t, from operating in its monopolized niche as in equation (2). The main difference emerges in competitive niches, where an obvious incumbent can troll entrants. The following expression for the value of trolling, identified with the superscript T, reflects our assumption on how trolls operate

$$v_t^T = e_t \alpha (1-\mu_1) v_t + \beta (1 - e_{t+1} [(1-\alpha) \lambda \mu_0 + \alpha]) v_{t+1}^T.$$

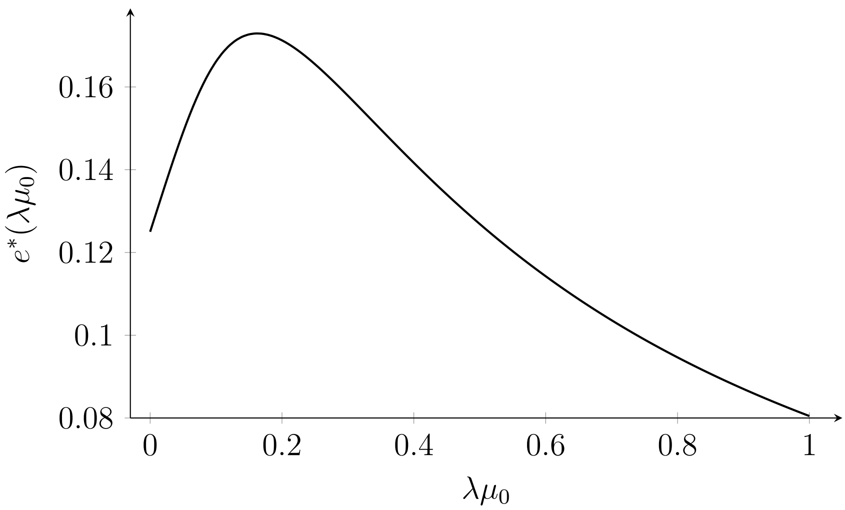
With probability e_t there is entry in the niche. Because of its patent right, the troll can challenge entrants in court. For a genuine entrant (with probability α), going to court means that the troll would succeed in blocking entry with probability $1 - \mu_1$. To prevent this outcome, the entrant negotiates a settlement with the troll. Assuming for simplicity that the entrant has all the bargaining power, the troll obtains a share $1 - \mu_1$ of all future rents from the innovation, v_t.[24] When the entrant is obvious (with probability $(1-\alpha)\lambda$) the troll sues and wins with probability $1 - \mu_0$, retaining the opportunity of extracting rents from a genuine entrant in the future. Otherwise, the obvious entrant becomes the new troll, and the market remains "competitive".

The proportion of monopolized niches, x_t, remains unchanged with respect to the baseline model in (3) since (i) the transition out of the monopolized state only occurs when an obvious entrant succeeds in court and (ii) whenever a genuine entrant reaches a "competitive" niche, its technology gets implemented, and the niche transitions to a monopolized state.

The new free entry condition involves a lower expected value from becoming a genuine incumbent and includes a new term that reflects the value of becoming a troll

$$\alpha \mu_1 v_t + (1-\alpha) \lambda \mu_0 v_t^T = 1.$$

[24]Results go through under any split of the rents from genuine innovation. Alternatively, we could assume that firms go to court and that when the patent of the entrant is invalidated, the troll takes possession of the technology and uses it to produce the upgraded product.



Note: Parameter values are $\alpha = 0.1$, $\beta = 0.8$, $\pi = 4.1$, and $\mu_1 = 0.5$.

Figure 12: Entry and Trolling. This figure shows that, even when trolling is allowed, entry is maximized at an interior value of the probability with which obvious innovators are allowed to enter, $\lambda\mu_0$.

Relative to the entry conditions resulting from the combination of (3) and (4) in the baseline model, the term in v_t does no longer include x_t. Genuine innovators now only expect to appropriate a share μ_1 of such a value, regardless of whether they enter a monopolized niche (where the incumbent blocks the entry with probability $1 - \mu_1$) of a "competitive" niche (where the troll appropriates a fraction $1 - \mu_1$ of the value via settlement). The term v_t^T reflects that the entrant is obvious with probability $(1 - \alpha)\lambda$ and is sued in every market, succeeding with probability μ_0 and becoming a troll. Although the proportion of monopolized niches now becomes irrelevant for entry, it still matters for welfare. The following result shows that patent trolling is detrimental to entry.

Lemma 9. *Compared to the benchmark model, when trolls can extract rents from future genuine innovators, steady-state entry is reduced.*

As Figure 12 shows, however, the forces behind the inverse U-shaped effect of the protection of genuine innovations against obvious entrants, $\lambda\mu_0$, on entry extend to this environment. Since the welfare expression (10) and the judge's problem (17) remain unchanged, the results of the baseline model are qualitatively preserved.

6 Concluding Remarks

Innovation is considered key to industry dynamics. Entry, exit, and innovation are complex interrelated phenomena in every industry, and especially so in those that are youngest and more technology intensive. Many of these industries rely on IPRs as the source of temporary monopoly power that allows the successful innovators to obtain a return from their previous R&D investments. Intellectual property, however, is a double-edged sword for the dynamics of innovative industries, as the protection of incumbent innovators may be an obstacle to the success of genuine innovators.

This paper contributes to the growing literature that analyzes the role of IPRs by embedding it in an industry-dynamics framework in which the value of innovation is stochastic and new firms replace existing incumbents. We find that innovation and welfare are maximized when incumbents receive maximal protection against small (obvious) improvements but as little as possible against large (genuine) innovations. However, if incumbents receive some protection against large innovations, allowing for some small innovations may be socially beneficial.

We also delve into the interaction between the mechanisms that implement the level of protection of incumbents against each kind of innovation, namely the ex-ante screening by the patent office and the ex-post enforcement by courts. The patent office can expend resources in trying to screen out small innovations before their very entry. Courts enter the game once an incumbent claims that its IPRs have been infringed by an entrant and can at that stage invest resources to improve the accuracy of their rulings. We show that courts provide better enforcement the more the patent office engages in the ex-ante screening. This complementarity shapes the choice of the socially optimal level of screening by the patent office and has significant policy implications for the design of the patent system. Insofar as standard cost-benefit analyses of the patent system neglect the indirect positive impact of patent screening on patent enforcement, they may underestimate the value of better patent screening.

Our analysis has focused on the impact of patent screening and enforcement on sequential innovation, uncovering new mechanisms regarding industry dynamics that influence the endogenous enforcement of patent rights by courts and engender the above-mentioned complementarity with ex-ante screening. Incentives to innovate in a sequential setup can also be affected by factors beyond patent screening that we have abstracted from. Extending our dynamic analysis with endogenous enforcement of other dimensions of the patent system such as patent breadth and patentability standards or the related topic of licensing could be interesting avenues for future research.

References

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, 1992, *60* (2), pp. 323–351.

Allison, John R. and Mark A. Lemley, "Empirical evidence on the validity of litigated patents," *AIPLA Quarterly Journal*, 1998, *26* (3), p. 185.

Atal, Vidya and Talia Bar, "Patent quality and a two-tiered patent system," *The Journal of Industrial Economics*, 2014, *62* (3), pp. 503–540.

Bessen, James and Eric Maskin, "Sequential Innovation, Patents, and Imitation," *Rand Journal of Economics*, 2009, *40* (4), pp. 611–635.

Caillaud, Bernard and Anne Duchêne, "Patent Office in innovation policy: Nobody's perfect," *International Journal of Industrial Organization*, 2011, *29* (2), pp. 242–252.

Choi, Jay Pil, "Patent Litigation as an Information-Transmission Mechanism," *American Economic Review*, 1998, *88* (5), pp. 1249–1263.

Daughety, Andrew F. and Jennifer F. Reinganum, "Keeping Society in the Dark: On the Admissibility of Pretrial Negotiations as Evidence in Court," *Rand Journal of Economics*, 1995, *26* (2), pp. 203–221.

Daughety, Andrew F and Jennifer F Reinganum, "Appealing judgments," *Rand Journal of Economics*, 2000, *31* (3), pp. 502–525.

—, "Informal sanctions on prosecutors and defendants and the disposition of criminal cases," *Journal of Law, Economics, and Organization*, 2016, *32* (2), pp. 359–394.

—, "Reducing Unjust Convictions: Plea Bargaining, Trial, and Evidence Disclosure," *Journal of Law, Economics, and Organization*, 2020, *36* (2), pp. 378–414.

Denicolò, Vincenzo, "Two-Stage Patent Races and Patent Policy," *Rand Journal of Economics*, 2000, *31* (3), pp. 488–501.

Denicolò, Vincenzo and Piercarlo Zanchettin, "How should forward patent protection be provided?" *International Journal of Industrial Organization*, 2002, *20* (6), pp. 801–827.

DENICOLÒ, VINCENZO AND PIERCARLO ZANCHETTIN, "What Causes Over-investment in R&D in Endogenous Growth Models?" *Economic Journal*, 2014, *124*(581), pp. 1192–1212.

DIXIT, AVINASH K. AND JOSEPH E. STIGLITZ, "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 1977, *67*(3), pp. 297–308.

FARRELL, JOSEPH AND ROBERT P. MERGES, "Incentives to challenge and defend patents: Why litigation won't help reliably fix patent office errors and administrative patent review might help," *Berkeley Technology Law Journal*, 2004, *19*(3), pp. 943–970.

FARRELL, JOSEPH AND CARL SHAPIRO, "How Strong Are Weak Patents," *American Economic Review*, 2008, *98*(4), pp. 1347–1369.

FRAKES, MICHAEL D. AND MELISSA F. WASSERMAN, "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data," *Review of Economics and Statistics*, 2017, *99*(3), pp. 550–563.

—, "Irrational Ignorance at the Patent Office," *Vanderbilt Law Review*, 2019, *72*(3), pp. 975–1030.

GALASSO, ALBERTO AND MARK SCHANKERMAN, "Patents and cumulative innovation: Causal evidence from the courts," *Quarterly Journal of Economics*, 2015, *130*(1), pp. 317–369.

GILBERT, RICHARD J AND DAVID M G NEWBERY, "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, 1982, *72*(3), pp. 514–526.

GREEN, JERRY R. AND SUZANNE SCOTCHMER, "On the Division of Profit in Sequential Innovation," *Rand Journal of Economics*, 1995, *26*(1), pp. 20–33.

GROSSMAN, GENE M AND ELHANAN HELPMAN, "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 1991, *58*(1), pp. 43–61.

GROSSMAN, GENE M AND MICHAEL L KATZ, "Plea bargaining and social welfare," *American Economic Review*, 1983, *73*(4), pp. 749–757.

HELLER, MICHAEL A. AND REBECCA S. EISENBERG, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science*, 1998, *280*, pp. 698–701.

Hopenhayn, Hugo, Gerard Llobet and Matthew Mitchell, "Rewarding Sequential Innovations: Prizes, Patents and Buyouts," *Journal of Political Economy*, 2006, *104*(6), pp. 1041–1068.

Landeo, Claudia M, Maxim Nikitin and Scott Baker, "Deterrence, lawsuits, and litigation outcomes under court errors," *Journal of Law, Economics, & Organization*, 2007, *23*(1), pp. 57–97.

Lee, Tom and Louis L Wilde, "Market Structure and Innovation: A Reformulation," *Quarterly Journal of Economics*, 1980, *94*(2), pp. 429–36.

Lei, Zhen and Brian D. Wright, "Why weak patents? Testing the examiner ignorance hypothesis," *Journal of Public Economics*, 2017, *148*, pp. 43–56.

Lemley, Mark A and Bhaven Sampat, "Is the patent office a rubber stamp," *Emory Law Journal*, 2008, *58*, p. 181.

Lemley, Mark A and Carl Shapiro, "Probabilistic patents," *Journal of Economic Perspectives*, 2005, *19*(2), pp. 75–98.

Lemley, Mark A. and Carl Shapiro, "Patent Holdup and Royalty Stacking," *Texas Law Review*, 2007, *85*, pp. 1991–2049.

Lemley, Mark D., "Rational Ignorance at the Patent Office," *Northwestern University Law Review*, 2001, *95*(1495-1532).

—, "Ignoring Patents," *Michigan State Law Review*, 2008, pp. 19–34.

Llobet, Gerard, "Patent Litigation When Innovation is Cumulative," *International Journal of Industrial Organization*, 2003, *21*(8), pp. 1135–1157.

Loury, Glenn C, "Market Structure and Innovation," *Quarterly Journal of Economics*, 1979, *93*(3), pp. 395–410.

Marshall, Guillermo and Álvaro Parra, "Innovation and competition: The role of the product market," *International Journal of Industrial Organization*, 2019, *65*, pp. 221–247.

O'Donoghue, Ted, "A Patentability Requirement for Sequential Innovation," *Rand Journal of Economics*, 1998, *29*(4), pp. pp. 654–679.

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

Parra, Álvaro, "Sequential innovation, patent policy, and the dynamics of the replacement effect," *Rand Journal of Economics*, 2019, *50*(3), pp. 568–590.

Picard, Pierre M. and Bruno van Pottelsberghe de la Potterie, "Patent office governance and patent examination quality," *Journal of Public Economics*, 2013, *104*, pp. 14–25.

Reinganum, Jennifer F, "Plea bargaining and prosecutorial discretion," *American Economic Review*, 1988, pp. 713–728.

Schankerman, Mark and Florian Schuett, "Patent Screening, Innovation, and Welfare," *The Review of Economic Studies*, 10 2021, *89*(4), pp. 2101–2148.

Scotchmer, Suzanne, *Innovation and Incentives*, MIT Press, Cambridge, MA, 2004.

Scotchmer, Suzanne and Jerry Green, "Novelty and Disclosure in Patent Law," *Rand Journal of Economics*, 1990, *21*(1), pp. pp. 131–146.

Segal, Ilya and Michael D. Whinston, "Antitrust in Innovative Industries," *American Economic Review*, 2007, *97*(5), pp. 1703–1730.

Spier, Kathryn E, "Pretrial bargaining and the design of fee-shifting rules," *Rand Journal of Economics*, 1994, pp. 197–214.

Spier, Kathryn E., "Litigation," in A. Mitchell Polinsky and Steven Shavell, eds., *Handbook of Law and Economics*, volume 1, Elsevier, 2007 pp. 259–342.

Appendix

A Proofs

Proof of Lemma 1. To show existence we start by imposing the steady condition $x_t = x$ for all t in eq. (3), obtaining eq. (6). Replacing x into (4) we obtain eq. (7). Using the free-entry condition (5), we obtain $v = p^{-1}$. Finally, using condition (5) again and in the steady state of (2), we find the expression for the entry flow (8). Because all these equations are linear, we have a unique solution.

To determine when the entry flow is positive, observe that (8) is positive if and only if $\pi > \underline{\pi}$ and lower than one if and only if $\pi < \underline{\pi} + \beta[\alpha + (1-\alpha)\lambda\mu_0]/\alpha$. ■

Proof of Lemmas 2 and 3. Most comparative statics are direct, as they follow from direct differentiation. The derivative of the entry flow with respect to $\lambda\mu_0$ is

$$\frac{de}{d\left(\lambda\mu_0\right)} = \frac{\left(1-\alpha\right)\left(1-\beta\right)}{\beta\left(\alpha\mu_1 + \left(1-\alpha\right)\lambda\mu_0\right)^2}\left[1 - \Pi\frac{p^2}{\alpha}\right] \tag{22}$$

where $\Pi = \pi/(1-\beta)$ is the present social value of a genuine innovation. The sign of the derivative is given by the sign of the term in square brackets. This term is a monotonically decreasing function of Π. To show that the derivative can be positive for any value of $\lambda\mu_0$, take the maximum value of Π for which there is no entry; i.e., take $\Pi = \underline{\pi}/(1-\beta) \equiv p^{-1}$. In this case, the bracket term becomes $1 - p/\alpha$, which is positive whenever $\mu_1 < 1$, as $p < \alpha$. By continuity, the results holds for values of $\Pi > p^{-1}$ for which there is entry.

To show that the derivative is negative when $\mu_1 = 1$, observe that in this case $p = \alpha$ and the parenthesis becomes $1 - \Pi\alpha$. In this scenario, Assumption 1 is equivalent to $\Pi > p^{-1} = \alpha^{-1}$ and the derivative is negative for any value of $\lambda\mu_0$. Finally, an example of an inverted-U shape relation is given in the main text. ■c

Proof of Proposition 1. For an interior solution of $\hat{\lambda}$ we compute the derivative (11) and solve for the values of λ such that (11) is equal to zero. Because the first order condition corresponds to a third degree polynomial, we obtain three candidate solution. Call them $\{0, -, +\}$. The first solution,

$$\lambda^0 = \frac{-\alpha\left(\Pi\alpha\mu_1 - 1\right)}{\mu_0\left(1-\alpha\right)\left(\Pi\alpha - 1\right)} < 0$$

where $\Pi = \pi/(1-\beta)$, corresponds to a value for which the welfare function is zero ($W(\lambda^0) = 0$). This occurs as $e(\lambda^0) = 0$ and $p(\lambda^0) = \Pi^{-1}$ hold. Using these facts, we show that the second order condition at λ^0 is positive. Start by observing

$$\frac{d^2W}{d\lambda^2} = \frac{d^2e}{d\lambda^2}\left(\Pi p - 1\right) + 2\frac{de}{d\lambda}\frac{dp}{d\lambda}\Pi + e\frac{d^2p}{d\lambda^2}\Pi = 2\frac{de}{d\lambda}\frac{dp}{d\lambda}\Pi$$

Proposition 2 implies that $dp/d\lambda > 0$. Similarly, using $p(\lambda^0) = \Pi^{-1}$ and Assumption 1 in (22) we can verify $de/d\lambda > 0$. Therefore, the critical point corresponding to a (local) minimum and a feasible maximum (i.e., with $\hat{\lambda} \in [0,1]$) must lie to the right of λ^0.

The second and third solutions are given by

$$\lambda^{-,+} = \lambda^0 + \frac{\alpha\,(1-\mu_1)\left(\Pi\alpha \pm \sqrt{\alpha\Pi\,(\Pi\alpha+8)}\right)}{2\mu_0\,(1-\alpha)\,(\Pi\alpha-1)}$$

Given Assumption 1, the denominator above is positive and the sign of the fraction depends on solution. For the negative solution the numerator is negative, and positive for the positive solution. Therefore, we have that $\lambda^- < \lambda^0 < \lambda^+$. Thus, given the analysis above, λ^+ is the only feasible solution. It can be readily verified, through differentiation, that λ^+ decreases in Π and that it is less than one whenever

$$\Pi \leq K \equiv \frac{(\alpha+(1-\alpha)\mu_0)^2}{\alpha(\alpha\mu_1+(1-\alpha)\mu_0)(\alpha(2\mu_1-1)+(1-\alpha)\mu_0)}.$$

Finally, it can be verified that λ^+ decreases in μ_1 and $\lambda^+ < 0$ when $\mu_1 = 1$. Consequently, there exists $\hat{\mu}_1$ such that, $\mu_1 > \hat{\mu}_1$ implies that $\hat{\lambda} = 0$, which proves the result. ■

Proof of Lemma 4. Subtracting the value functions w_M and w_C we obtain

$$w_M - w_C = -e\alpha\Pi\,(1-\mu_1) + [1 - e\,(\alpha+(1-\alpha)\lambda\mu_0)]\,\beta(w_M - w_C),$$

Solving for $w_M - w_C$ delivers (14). ■

Proof of Lemma 5. Fix a triplet $(s,\, \hat{s},\, \gamma)$. Using $E_I = \Pi$, E_{II} given by equation (14), and equation (1) we can rewrite $J\,(s,\hat{s},\gamma)$ as:

$$\frac{\Pi}{2}\,(1-\gamma)\,(1-s)\underbrace{\left(1 - \frac{\gamma\alpha\beta\,(1-\mu_1\,(\hat{s}))\,e\,(\hat{s},\gamma)}{(1-\beta)\,(1-\gamma)+\beta\alpha\,((1-\gamma)+\gamma\mu_0\,(\hat{s}))\,e\,(\hat{s},\gamma)}\right)}_{\Phi(\hat{s},\gamma)} + c\,(s)\,.$$

Working the parenthesis and using that $\mu_0(s)+\mu_1(s) = 1$ for all s, we obtain the expression in (18). Differentiating $\Phi(\hat{s},\gamma)$ with respect to $e\,(\hat{s},\gamma)$ we obtain

$$\frac{d\Phi(\hat{s},\gamma)}{de} = -\frac{\alpha\beta\,(1-\beta)\,(1-\gamma)\,\gamma\mu_0(\hat{s})}{[(1-\beta)\,(1-\gamma)+\beta\alpha\,((1-\gamma)+\gamma\mu_0(\hat{s}))\,e\,(\hat{s},\gamma)]^2} < 0.$$

Because it will be useful for the comparative statics and to prove other propositions, we replace $e\,(\hat{s},\gamma)$ (see equation (8)) into $\Phi(\hat{s},\gamma)$ to obtain

$$\Phi\,(\hat{s},\gamma) = \frac{(1-\gamma)\,[\Pi\alpha k_1 - k_2] + k_1 k_2}{k_2\,(\Pi\alpha k_1 - (1-\mu_1(\hat{s}))\,(1-\gamma))} \tag{23}$$

where $k_1 = (1-\gamma)\,\mu_1(\hat{s}) + \gamma\mu_0(\hat{s})$ and $k_2 = 1-\gamma+\gamma\mu_0(\hat{s})$ are functions of $\hat{s}$ and γ. We omit this dependence for parsimonious notation. Observe that, for a given γ, α and Π are always together in (23). Thus,

$$\frac{d\Phi(\hat{s},\gamma)}{d\Pi\alpha} = -\frac{\gamma\mu_0(\hat{s})k_1^2}{k_2(\Pi\alpha k_1-(1-\mu_1(\hat{s}))(1-\gamma))^2} < 0$$

which proves the result for α and Π. Differentiating (23) with respect to γ we obtain

$$\frac{d\Phi\,(\hat{s},\gamma)}{d\gamma} = -\mu_0\frac{\Pi\alpha k_1\,(\Pi\alpha k_1 - [k_2 + (1-\mu_1(\hat{s}))\,(1-\gamma)]) + k_2^2(1-\mu_1(\hat{s}))}{k_2^2\,(\Pi\alpha k_1 - (1-\mu_1(\hat{s}))\,(1-\gamma))^2}.$$

Observe that the denominator is always positive. Assumption 1 implies $\Pi > k_2/(\alpha k_1) = p^{-1}$. Because the numerator is increasing in Π, we can replace the minimal feasible profit in the numerator to obtain a upper bound for the derivative

$$\frac{d\Phi\left(\gamma, \hat{s}\right)}{d\gamma} < -\frac{\gamma\mu_0^2\left(1-\mu_1(\hat{s})\right)}{k_2\left(\Pi\alpha k_1 - \left(1-\mu_1(\hat{s})\right)\left(1-\gamma\right)\right)^2} \leq 0$$

which proves the last comparative statics result. ■

Proof of Proposition 2. To show existence of $\gamma^*(\hat{s})$ simply observe that, when $\gamma = 0$, $J(0, \hat{s}, 0) = \Pi/2$. Also, when $\gamma = 1$, $J(0, \hat{s}, 1) = 0$. Because J is continuous in γ, if $c < \Pi/2$, Intermediate Value Theorem implies there exists $\gamma^*(\hat{s}) \in (0,1)$ such that $J(0, \hat{s}, \gamma^*(\hat{s})) = c$. To prove that $\gamma^*(\hat{s})$ is unique, we show that $J(0, \hat{s}, \gamma)$ is decreasing in γ. Computing the derivative of $J(0, \hat{s}, \gamma)$ with respect to γ we obtain

$$\frac{dJ\left(0, \hat{s}, \gamma\right)}{d\gamma} = \frac{\Pi}{2}\left(-\Phi\left(\hat{s}, \gamma\right) + \left(1-\gamma\right)\frac{d\Phi\left(\hat{s}, \gamma\right)}{d\gamma}\right),$$

which, by Lemma 5, is negative. Thus, $dJ/d\gamma < 0$ and uniqueness follows. If $c \geq \Pi/2$, J decreasing in γ implies that the cost of effort is always higher than no exerting effort. ■

Proof of Lemma 6. It can be readily verified that, when $\hat{s} = 1$, $\Phi(1, \gamma) = 1$. Consequently, solving for $J(0, 1, \gamma) = c$ we obtain that $\gamma^*(1) = (\Pi - 2c)/\Pi$. Also, when $\hat{s} = 0$, $\Phi(0, \gamma) < 1$ for all $\gamma > 0$. Thus, $J(0, 1, \gamma) > J(0, 0, \gamma)$ for all $\gamma \in (0, 1)$. Starting from $\gamma^*(0) \in (0, 1)$. The previous observation implies $c = J(0, 0, \gamma^*(0)) < J(0, 1, \gamma^*(0))$. Then, because J is decreasing in γ (see Proposition 2), $\gamma^*(0) < \gamma^*(1)$. ■

Proof or Proposition 3. The characterization of the equilibrium is immediate from Lemma 6 and the discussion in the text. Regarding the equilibrium entry, observe

$$e\left(0, \gamma\right) = \frac{2\left(1-\gamma\right)\left(1-\beta\right)}{\alpha\beta}\left(\frac{\Pi\alpha}{2-\gamma} - 1\right) \qquad e\left(1, \gamma\right) = \frac{1-\beta}{\alpha\beta}\left(\Pi\alpha - 1\right).$$

If $\Pi \leq (2-\gamma)/\alpha$, then $e(0, \gamma) = 0$ and $e(1, \gamma) \geq 0$. If $\Pi > (2-\gamma)/\alpha$ then $e(0, \gamma) > 0$ and $e(1) > e(0)$ if and only if

$$\Pi > \frac{\left(2-\gamma\right)}{\alpha}\frac{\left(2\gamma-1\right)}{\gamma}.$$

Since $(2\gamma - 1)/\gamma < 1$ for $\gamma < 1$ and, in equilibrium, $\gamma^*(\hat{s}) < 1$, the result holds. ■

Proof of Proposition 4. As explained in the main text the result arises from the complementarity of the decisions of individual judges. Hence, to conduct the comparative statics exercise we only need to show that the effort level of a judge is monotonically decreasing in c and α and increasing in Π. The effect c is immediate, higher effort costs decreases, on the margin, a judge's effort. Regarding the effects of an increase in Π and α, we need to compute the sign of

$$\frac{dJ}{d\Pi} = \frac{1-\gamma}{2}\left(\Phi + \Pi\frac{d\Phi}{d\Pi}\right) \text{ and } \frac{dJ}{d\alpha} = \frac{\Pi(1-\gamma)}{2}\frac{\partial\Phi}{\partial\alpha},$$

respectively. From Lemma 5 we know that $\Phi(\hat{s},\gamma)$ is decreasing in α for all $\hat{s}$ which is enough to show the effect of this parameter. For Π we have that

$$\frac{dJ}{d\Pi}=\frac{1-\gamma}{2}\frac{\left[(1-\gamma)\left(\Pi\alpha k_1-k_2\right)+k_1k_2\right]\left[\alpha\Pi k_1-(1-\mu_1(\hat{s}))(1-\gamma)\right]-\gamma\mu_0(\hat{s})\Pi\alpha k_1^2}{k_2(\Pi\alpha k_1-(1-\mu_1(\hat{s}))(1-\gamma))^2},$$

where k_1 and k_2 are defined in Lemma 5. We show that a lower bound of this derivative is positive. Notice that positive entry (assumption 1) implies that $\Pi > k_2/(\alpha k_1)$. As a result both $\Pi\alpha k_1-k_2>0$ and $\Pi\alpha k_1-(1-\mu_1(\hat{s}))(1-\gamma)>0$. This means that a lower bound for the first square bracket is k_1k_2. Substituting and rearranging we obtain

$$\frac{dJ}{d\Pi}>\frac{(1-\gamma)\left[\Pi\alpha k_1^2-(1-\mu_1(\hat{s}))k_1k_2\right]}{k_2(\Pi\alpha k_1-(1-\mu_1(\hat{s}))(1-\gamma))^2}>\frac{(1-\gamma)\mu_1(\hat{s})k_1k_2}{k_2(\Pi\alpha k_1-(1-\mu_1(\hat{s}))(1-\gamma))^2}>0,$$

where we used $k_2/(\alpha k_1)$ as a lower bound for Π, proving the result. ■

Proof of Proposition 5. For a given c, we want to solve $\lambda^*(c)\in\arg\max W^{sp}(\lambda,c)$ where, using equation (19),

$$W^{sp}(\lambda,c)=\begin{cases}W(\lambda;1,c) & \text{if } \lambda\leq\lambda_J(c)\\ W(\lambda;0) & \text{if } \lambda>\lambda_J(c)\end{cases}\quad\text{and}\quad\lambda_J(c)=\frac{\alpha}{1-\alpha}\frac{\Pi-2c}{2c}\tag{24}$$

is the minimum screening quality that induces enforcement effort by the judges; $\lambda_J(c)$ comes from replacing $\gamma^*(1)=(\Pi-2c)/\Pi$ into equation (15). This function is decreasing in c, takes the value of zero when $c=\Pi/2$ and diverges when the enforcement cost approaches zero (see Figure 13).

We aim to find $\lambda^*(c)$ and study how it changes with the cost of screening c. To start, we characterize the welfare function under no enforcement effort.

Lemma A.1. *Under Assumption 3, the function $W(\lambda;0)$ is single peaked when $\lambda\in[0,1]$.*

Proof. To ease the notation, redefine $e(\lambda)=e(0,\lambda)$ and $p(\lambda)=p(0,\lambda)$. Recall $W(\lambda;0)=e(\lambda)[\Pi p(\lambda)-1-\kappa(\lambda)]$ and observe that $W(1;0)>0$ and $\lim_{\lambda\to 0}W(\lambda,0)=-\infty$. We show that there is a unique $\lambda_0\in(0,1]$ in the $\arg\max W(\lambda;0)$ satisfying $W(\lambda_0;0)>0$. The first-order condition (FOC) for an interior maximum is given by

$$0=e'(\lambda_0)\left(\Pi p(\lambda_0)-1-\kappa(\lambda_0)\right)+e(\lambda_0)\left(\Pi p'(\lambda_0)-\kappa'(\lambda_0)\right).$$

Observe that $e(\lambda),-\kappa'(\lambda),p'(\lambda)>0$ by Assumptions 1 and 3, and Lemma 2. Thus, $e(\lambda_0)[\Pi p'(\lambda)-\kappa'(\lambda)]>0$ for all λ. Also, because $W(1;0)>0$, if an interior solution for λ_0 exists, it satisfies $W(\lambda_0;0)>0$, implying $\Pi p(\lambda_0)-1-\kappa(\lambda_0)>0$ and $e'(\lambda_0)<0$.

We show that, every solution to the FOC satisfies the second-order conditions for an interior maximum, implying that there is at most one interior maximum. In the two feasible cases $\lambda_0\in(0,1)$ or $\lambda_0=1$, the function $W(\lambda;0)>0$ is single peaked.

The second-order condition is given by

$$e''(\lambda_0)\left(\Pi p(\lambda_0)-1-\kappa(\lambda_0)\right)+2e'(\lambda_0)\left(\Pi p'(\lambda_0)-\kappa'(\lambda_0)\right)+e(\lambda_0)\left(\Pi p''(\lambda_0)-\kappa''(\lambda_0)\right).$$

Substituting in $\Pi p(\lambda_0)-1-\kappa(\lambda_0)=e(\lambda_0)\left[\Pi p'(\lambda_0)-\kappa'(\lambda_0)\right]/[-e'(\lambda_0)]$ from the FOC, we obtain

$$\left(e''(\lambda_0)e(\lambda_0)-2e''(\lambda_0)^2\right)\left[\Pi p'(\lambda_0)-\kappa'(\lambda_0)\right]\left[-e'(\lambda_0)\right]^{-1}+e(\lambda_0)\left[\Pi p''(\lambda_0)-\kappa''(\lambda_0)\right],$$

which is negative, as $\Pi p''(\lambda)-\kappa''(\lambda)<0$ and $e''(\lambda)e(\lambda)-2e''(\lambda)^2<0$ for $\lambda>0$. Therefore, every point satisfying the FOC is a maximum and, at most, one maximum exist. Thus, the function is single peaked in $\lambda \in [0,1]$ (where λ_0 is either interior or 1). ■

We now turn to examine the scenario when judges exert full-enforcement effort; that is, we study the $\arg\max W(\lambda;1,c)$, which has a unique solution, called $\lambda_1(c)$.

Lemma A.2. *Under Assumption 3 the functions $\lambda_1(c)$ and $\lambda_J(c)$ single cross.*

Proof. The function $\lambda_1(c)$ is implicitly given by the first-order condition $-\kappa'(\lambda_1(c)) = (1-\alpha)c$; is decreasing in c; and, given the Inada conditions, it takes the value of 1 when $c=0$ and approaches zero from above as c goes to infinity. These observations imply that the functions $\lambda_1(c)$ and $\lambda_J(c)$ cross at least once. Implicit differentiation give us $\lambda_1'(c) = -(1-\alpha)/\kappa''(\lambda_1(c)) < 0$. On the other hand $\lambda_J'(c) = -\alpha\Pi/[2(1-\alpha)c^2]$. Then, under Assumption 3, $\lambda_J'(c) < \lambda_1'(c) < 0$ for $c \leq \Pi/2$, and the functions cross once. ■

We define c_1 to be the unique cost satisfying $\lambda_1(c_1) = \lambda_J(c_1)$. The previous analysis implies that $\lambda_1(c) < \lambda_J(c)$ for all $c < c_1$. Similarly, define c_0 to be the unique cost satisfying $\lambda_J(c_0) = \lambda_0$. The cost c_0 can be above or below c_1 (see Figure 13).

Lemma A.3. *The solution to* $\arg\max W(\lambda;1,c)$ *subject to* $\lambda \leq \lambda_J(c)$ *and* $\arg\max W(\lambda;0)$ *subject to* $\lambda \geq \lambda_J(c)$ *is given by* $\lambda_1^-(c) = \min\{\lambda_1(c),\lambda_J(c)\}$ *and* $\lambda_0^+(c) = \max\{\lambda_0,\lambda_J(c)\}$.

Proof. Because $W(\lambda;1,c)$ and $W(\lambda;0)$ are single peaked in λ, the solution when the constraint binds is $\lambda_J(c)$. ■

Lemma A.4. *The function $W(\lambda_1(c);1,c)$ is strictly decreasing in c, $W(\lambda_J(c);1,c)$ is strictly decreasing in c whenever $c \geq c_1$, and $W(\lambda_J(c);0)$ is increasing in c for $c > c_0$.*

Proof. The derivative of $W(\lambda_1(c);1,c)$, using the Envelope Theorem, becomes

$$\frac{dW}{dc} = -(\alpha + (1-\alpha)\lambda_1(c)) < 0.$$

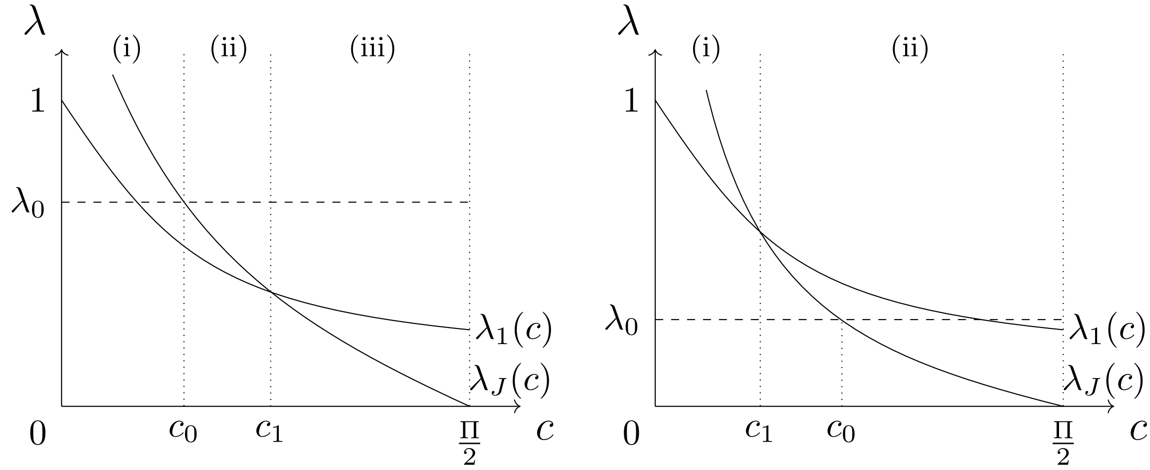
The derivative of $W(\lambda_J(c);1,c)$ with respect c can be written as

$$\frac{dW}{dc} = -\left(\kappa'(\lambda_J(c)) + c(1-\alpha)\right)\frac{\partial\lambda_J(c)}{\partial c} - (\alpha + (1-\alpha)\lambda_J(c)).$$

Because $c \geq c_1$, we have that $\lambda_J(c) \leq \lambda_1(c)$ and, consequently, $\kappa'(\lambda_J(c)) \leq \kappa'(\lambda_1(c)) \leq 0$ Then, $\kappa'(\lambda_J(c)) + c(1-\alpha) \leq \kappa'(\lambda_1(c)) + c(1-\alpha)) = 0$ and the result follows from observing $\partial\lambda_J/\partial c < 0$. Finally, because $W(\lambda,0)$ is single peaked in λ and $\lambda_J(c)$ for $c \leq c_0$ converges to λ_0 from above, $W(\lambda_J(c),0)$ increases to the maximum. ■

Lemma A.5. *There is a unique cutoff $c_{10} > 0$ satisfying: $W(\lambda_1(c_{10});1,c_{10}) = W(\lambda_0;0)$. For costs below the cutoff (for $c < c_{10}$) it satisfies: $W(\lambda_1(c);1,c) > W(\lambda_0;0)$ and the opposite inequality for enforcement costs higher than the cutoff.*

Proof. It can be readily verified that, when $c=0$, $W(\lambda_1(0);1,0) > W(\lambda_0;0)$. By Lemma A.4 the function $W(\lambda_1(c);1,c)$ is strictly decreasing in c and $\lim_{c\to\infty} W(\lambda_1(c);1,c) = -\infty$, there is a unique value of c, call it c_{10}, such that $W(\lambda_1(c_{10});1,c_{10}) = W(\lambda_0;0)$. The inequality follow from $W(\lambda_1(c);1,c)$ being decreasing. ■



(a) Scenario $c_0 \leq c_1$ **(b)** Scenario $c_1 < c_0$

Figure 13: The full enforcement effort constraint.

To solve for the socially optimal screening rate, we need to study two scenarios: $c_0 \leq c_1$, and $c_1 < c_0$. The analysis of these two scenarios is divided into sub cases. For each scenario below, it is recommended to follow the proof along with Figure 13.

Scenario $c_0 \leq c_1$ *case* (i) $c_{10} \leq c_0$: By definition of c_{10}, for any $c \leq c_{10}$ and under optimal unconstrained screening, full judiciary effort is socially preferred to no effort. Also, by $c \leq c_1$, full judiciary effort under unconstrained screening is feasible. Thus, $\lambda^*(c) = \lambda_1(c)$ in $c \leq c_{10}$. For $c \in (c_{10}, c_0]$, no judiciary effort under unconstrained screening is preferred over full effort but it is unfeasible as $\lambda_0 \leq \lambda_j(c)$. The patent office needs to compare full judiciary effort under unconstrained screening versus no-effort under constrained screening; that is, $W(\lambda_1(c); 1, c) \gtrless W(\lambda_J(c); 0)$. Observe that $W(\lambda_1(c_{10}); 1, c_{10}) = W(\lambda_0; 0) > W(\lambda_J(c_{10}); 0)$ and $W(\lambda_1(c_0); 1, c_0) < W(\lambda_0; 0) = W(\lambda_J(c_0); 0)$. Thus, by the intermediate value theorem, there exists $c_{1J} \in (c_{10}, c_0]$ such that $W(\lambda_1(c_{1J}); 1, c_{1J}) = W(\lambda_J(c_{1J}); 0)$. This value is unique because, by Lemma A.4, $W(\lambda_1(c); 1, c)$ decreases and $W(\lambda_J(c); 0)$ increases in c. Let $c^* = c_{1J}$. Then, when $c \leq c_{1J}$, we have $\lambda^*(c) = \lambda_1(c)$. At c^*, $\lambda^*(c)$ jumps up to $\lambda_J(c)$ which decreases in $(c_{1J}, c_0]$ to λ_0, where it remains constant for $c > c_0$.

Scenario $c_0 \leq c_1$ *case* (ii) $c_{10} \in (c_0, c_1]$: As before, for any $c \leq c_{10}$ full judiciary effort with unconstrained optimal screening is preferred over no effort. Also, by $c \leq c_1$, full judiciary effort under unconstrained screening is feasible and the optimal screening is $\lambda^*(c) = \lambda_1(c)$. Let $c^* = c_{10}$. When $c > c_{10}$, no judiciary effort with unconstrained screening is preferred and is feasible as $c > c_0$; $\lambda^*(c)$ jumps upwards to λ_0 for $c > c_{10}$.

Scenario $c_0 \leq c_1$ *case* (iii) $c_{10} > c_1$: For $c \leq c_1$, $\lambda^*(c) = \lambda_1(c)$ for the same reasons as previous cases. Then, when $c \in (c_1, c_{10}]$, unconstrained judiciary effort is ideal but infeasible. The patent office needs to compare judiciary effort with constrained screening versus no judiciary effort with unconstrained screening; i.e., $W(\lambda_J(c); 1, c) \gtrless W(\lambda_0; 0)$. By the intermediate value theorem, there exists a cost c_{J0} such that $W(\lambda_J(c_{J0}); 1, c_{J0}) = W(\lambda_0; 0)$.[25] The cost c_{J0} is unique because, by Lemma A.4, $W(\lambda_J(c); 1, c)$ decreases in c for $c > c_1$ and $W(\lambda_0; 0)$ is constant. Let $c^* = c_{J0}$. Then, $\lambda^*(c) = \lambda_J(c)$ in $(c_1, c_{J0}]$ because

[25]To see this observe: (a) $W(\lambda_J(c_1); 1, c_1) = W(\lambda_1(c_1); 1, c_1) > W(\lambda_0; 0)$ by $\lambda_J(c_1) = \lambda_1(c_1)$ and

the constrained screening under full judiciary effort delivers more welfare than no effort. At c_{J0}, $\lambda^*(c)$ jumps upward to λ_0 and remains there for $c > c_{J0}$.

Scenario $c_1 < c_0$ *case* (i) $c_{10} \leq c_1$: For $c \leq c_{10}$, $\lambda^*(c) = \lambda_1(c)$ for the same reasons as previous cases. For $c > c_{10}$, no judiciary effort under unconstrained screening is optimal, but only becomes feasible starting at c_0. Recall $\lambda_1^-(c) = \min\{\lambda_1(c), \lambda_J(c)\}$. In the interval $c \in (c_{10}, c_0)$, we need to compare the welfare of judiciary effort with potentially constrained screening $W(\lambda_1^-(c); 1, c)$—which is decreasing in c by Lemma A.4—to the welfare of no judiciary effort constrained $W(\lambda_J(c); 0)$—which increases in c. By the monotonicity of functions and the intermediate value theorem, there exists a unique cost $c_{JJ} \in (c_{10}, c_0)$ such that $W(\lambda_1^-(c_{JJ}); 1, c_{JJ}) = W(\lambda_J(c_{JJ}); 0)$. Let $c^* = c_{JJ}$. We have two scenarios:
(a) $c_{JJ} < c_1 \in (c_{10}, c_0)$: then, we have judiciary effort with unconstrained screening for $c \in (c_{10}, c_{JJ}]$. That is, $\lambda^*(c) = \lambda_1(c)$. For $c > c_{JJ}$ we have no judiciary effort with constrained screening. The optimal screening jumps up to $\lambda^*(c) = \lambda_J(c)$ and decreases smoothly to λ_0 when c_0, remaining constant for higher costs.
(b) $c_{JJ} > c_1 \in (c_{10}, c_0)$: then, for $c \in (c_{10}, c_1]$ we have judiciary effort with unconstrained screening, $\lambda^*(c) = \lambda_1(c)$. For $c \in (c_1, c_{JJ}]$ judiciary effort with constrained screening is optimal, $\lambda^*(c) = \lambda_J(c)$. For $c > c_{JJ}$ we have no judiciary effort with constrained screening, $\lambda^*(c) = \lambda_J(c)$, decreasing smoothly to λ_0 when c_0, remaining constant for higher costs.

Scenario $c_1 \leq c_0$ *case* (ii) $c_{10} > c_1$: Since $c_{10} > c_1$ we have that, for any $c \leq c_1$, $\lambda^*(c) = \lambda_1(c)$ for the same reasons as previous cases. Recall $\lambda_0^+(c) = \max\{\lambda_0, \lambda_J(c)\}$. Observe $W(\lambda_J(c_1); 1, c_1) = W(\lambda_1(c_1); 1, c_1) > W(\lambda_0^+(c), 0)$ (by $c_1 < c_{10}$) and $W(\lambda_J(c_{10}); 1, c_{10}) < W(\lambda_1(c_{10}); 1, c_{10}) = W(\lambda_0^+(c), 0)$. Then, by the intermediate value theorem, there exists a cost $c_{J0} \in (c_1, c_{10})$ such that $W(\lambda_J^*(c_{J0}); 1, c_{J0}) = W(\lambda_0^+(c_{J0}); 0)$. This cost is unique because, by Lemma A.4, the first function decreases and the second increases in c. Let $c^* = c_{J0}$. Observe that $c_{10} \gtrless c_0$. We have two scenarios:
(a) $c_{J0} \in (c_1, c_0]$. Then, the optimal screening is $\lambda^*(c) = \lambda_J(c)$, but the judiciary effort changes from full (when $c \in (c_1, c_{J0}]$) to no effort (when $c \in (c_{J0}, c_0]$). The optimal screening smoothly decreases to λ_0 reaching that point at $c = c_0$, remaining constant for higher costs.
(b) $c_{J0} \in (c_0, c_{10}]$. Then, for $c \in (c_1, c_{J0}]$ the optimal screening is $\lambda^*(c) = \lambda_J(c)$ with judges performing full effort. For $c > c_{J0}$ judges perform no effort and the optimal screening jumps to λ_0, remaining constant for higher costs. ■

Proof of Lemma 7. Let $m > 1$ be an arbitrary number of innovations in the quality ladder and $z_m = \delta^m z$ for any $z > 0$. In a monopolized niche, due to the unit elasticity of demand, the incumbent wants to charge the highest price feasible. In this case $p = z_{m-1}$. Then, the incumbent profits are given by $\pi = (p - z_m)q = a(1 - \delta)$ which is independent of the number of innovations and the baseline cost z. The dead-weight loss in the market is given by

$$\int_{z_m}^{z_{m-1}} q(p)dp - \pi = a\left(\ln\left(\delta^{-1}\right) - (1 - \delta)\right)$$

also independent of m and z. ■

Proof of Proposition 6. Because $s^{CS}(\hat{s}, \gamma) = s^*(\hat{s}, \gamma) + (1 - \gamma)L\Delta(\hat{s}, \gamma)/2\tilde{c}$ and $L > 0$, whether s^{CS} is higher than s^* depends on the sign of $\Delta(\hat{s}, \gamma)$. It is easy to verify that

$c_1 < c_{10}$, and; (b) $W(\lambda_J(c_{10}); 1, c_{10}) < W(\lambda_1(c_{10}); 1, c_{10}) = W(\lambda_0; 0)$. By continuity of $W(\lambda_J(c); 1, c)$ the result follows.

$\Delta(\hat{s},0) = 1$ and $\lim_{\gamma\to 1}\Delta(\hat{s},\gamma) = -\infty$. Because $\Delta(\hat{s},\gamma)$ is strictly decreasing in γ and is continuous, $\Delta(\hat{s},\gamma^{\circ}(\hat{s})) = 0$ exists and is uniquely defined. To show that $\gamma^{\circ}(\hat{s})$ is increasing in $\Pi\alpha$, observe that

$$\frac{d\Delta}{d\Pi\alpha} = \frac{\gamma k_1^2[(1-\gamma)(1-\mu_0(\hat{s}))+\gamma\mu_0(\hat{s})]}{(1-\gamma)k_2\left[\Pi\alpha k_1-(1-\mu_1(\hat{s})))(1-\gamma)\right]^2} > 0.$$

Thus, because $\Delta(\hat{s},\gamma)$ is decreasing in γ, higher $\Pi\alpha$ implies a higher $\gamma^{\circ}(\hat{s})$. ■

Proof of Proposition 7. Differentiating $s^*(\hat{s},\gamma)$ with respect to $\hat{s}$ we obtain

$$\frac{d\Phi}{ds} = \frac{2\gamma\,(1-\gamma)\,\hat{k}_1\left(\hat{k}_1+s+2\gamma-3\right)}{\hat{k}_2^2\left(\hat{k}_1-(1-\gamma)\,(1-s)\right)^2} \tag{25}$$

where $\hat{k}_1 = \Pi\alpha(1+s\,(1-2\gamma)) > 0$ and $\hat{k}_2 = 2-\gamma\,(1+s) > 0$ are the analogous expressions for k_1 and k_2 in Lemma 5 after replacing for the judge's beliefs using (1). To ease out notation, $\hat{k}_1$ also includes $\Pi\alpha$. The sign of this derivative is given by the term in parenthesis in the numerator, which is positive if and only if condition (20) holds. When $\hat{s}=1$ or $\gamma=1$ condition (20) becomes equal to $\Pi > p(\hat{s},\gamma)^{-1}$ which is equivalent to the positive entry assumption (Assumption 1). ■

Proof of Lemma 8. The social value of judge from not exerting effort in (16) once we replace $E_I = \Pi$ and E_{II} using equation (14) becomes[26]

$$J = \Pi\,(1-\gamma)\,(1-m)\,\Phi\,(\hat{s},m,\gamma)$$

where

$$\Phi\,(\hat{s},m,\gamma) = \frac{(1-\gamma)\left[(1-\beta)+\alpha\beta e\,(\hat{s},m,\gamma)\right]}{(1-\beta)\,(1-\gamma)+\beta\alpha\left((1-\gamma)+\gamma\mu_0(\hat{s},m)\right)e\,(\hat{s},m,\gamma)} \in [0,1],$$

and $e\,(\hat{s},m,\gamma)$ is given by (8) and its dependence on m is solely determined by (21). The proof that $\Phi\,(\hat{s},m,\gamma)$ decreases in γ is thus given by Lemma 5, as the result holds for generic μ_0 and μ_1. ■

Proof of Proposition 8. Solving the system of equations that determine the value of a niche in each of the states we obtain that $w_M - w_I = -e\Pi k_1/(1-\beta+\beta e k_1)$ where $k_1 \equiv \alpha(2-\mu_1)+(1-\alpha)\lambda\mu_0$. Computing the type-I error,

$$E_I = \Pi + \beta(w_M - w_I) = \frac{\Pi(1-\beta)}{1-\beta+\beta e k_1} > 0. \tag{26}$$

Notice that $e>0$ implies $E_I < \Pi$. Let $k_2 \equiv \alpha+(1-\alpha)\lambda\mu_0$, computing the type-II error we obtain

$$E_{II} = \beta(w_M - w_C) = -\frac{\Pi e\alpha(1-\beta)(1-\mu_1)}{(1-\beta+\beta e k_1)(1-\beta+\beta e k_2)} \leq 0, \tag{27}$$

[26] We used the following property in the derivation:

$$(1-m)\left((1-\gamma)+\gamma\mu_0\right)-\gamma m\,(1-\mu_1) = (1-\gamma)\,(1-m)\,.$$

which is strictly negative when $e > 0$ and $\mu_1 < 1$.

That $J(1, \hat{s}, \gamma) = c$ is immediate as it implies $\mu_1 = 1$ and $\mu_0 = 0$. Regarding $J(0, \hat{s}, \gamma) > 0$, take the expressions for E_I and E_{II} in (26) and (27). Using the expression (15) for λ in the numerator we can compute

$$J(0, \hat{s}, \gamma) = \frac{1-\gamma}{2} E_I + \frac{\gamma}{2} E_{II} = \frac{\pi}{2} \frac{(1-\beta)(1-\gamma) + e\alpha\left(\beta\left(1-\gamma+\gamma\mu_0\right) - \gamma(1-\mu_1)\right)}{(1-\beta+\beta e k_1)(1-\beta+\beta e k_2)},$$

where we used $\Pi(1-\beta) = \pi$. For a given entry level, the numerator is decreasing in γ. Then, it is enough to show that this expression is positive when evaluated at $\gamma^{\max} = 1-\alpha$. In that case, the numerator becomes

$$(1-\beta)\alpha + \alpha e(\beta\alpha - (1-\beta)(1-\alpha)\mu_0),$$

where we used the fact that $\mu_0 + \mu_1 = 1$. Since $e \leq 1$ the previous expression is always positive, indicating that $J(0, \hat{s}, \gamma) > 0$ for all values of γ. ■

Proof of Lemma 9. Solving for steady state for the value of a genuine incumbent and a the value of a troll we obtain

$$v = \frac{\pi}{1-\beta+e\beta\left[(1-\alpha)\lambda\mu_0 + \alpha\mu_1\right]} \quad \text{and} \quad v^T = \frac{e\alpha\ (1-\mu_1)\,v}{1-\beta+e\beta\left[(1-\alpha)\lambda\mu_0+\alpha\right]}.$$

Replacing the steady-state values in the free-entry condition we obtain

$$v\left[\alpha\mu_1 + (1-\alpha)\lambda\mu_0\left(\frac{e\alpha\,(1-\mu_1)}{1-\beta+e\beta\left[(1-\alpha)\lambda\mu_0+\alpha\right]}\right)\right] = 1.$$

This entry condition contrasts with the benchmark model where, in an interior solution, we have that $vp = 1$ where p is given by (7). Notice that v decreases in e and equals v in (9) when e takes its benchmark value in (8). Hence, to show that entry is lower under trolling, it is enough to show that the probability of entry rate in the benchmark model, p, is greater than the expression in square brackets. That is,

$$\alpha\frac{\alpha\mu_1 + (1-\alpha)\lambda\mu_0}{\alpha+(1-\alpha)\lambda\mu_0} > \alpha\mu_1 + (1-\alpha)\lambda\mu_0\left(\frac{e\alpha\,(1-\mu_1)}{1-\beta+e\beta\left[(1-\alpha)\lambda\mu_0+\alpha\right]}\right).$$

As it turns out, this is true for any value of $e \in [0, 1]$. ■

B Extensions: Technical Details

B.1 Cost-saving Innovations and Static Inefficiency

In this appendix we present the details of the cost-saving innovation model introduced in Section 5.1. We start deriving the objective function for the patent office and then move to the judges' problem.

Social Welfare Because profits are invariant to the cumulative number of innovations, the firm behavior and industry dynamics described in Section 3 go through without alterations. The objective functions of the social planner and courts, however, need to be reformulated to account for the deadweight loss ℓ. Figure 5 depicts the product market payoff associated with each innovation. Building on the expression for the deadweight losses obtained in Lemma 7, social welfare can be written as follows

$$W = e(\hat{s}, \gamma(\lambda)) \left[(1-\alpha)\lambda x \mu_0(\hat{s}) L + \alpha \big(x\mu_1(\hat{s})(\Pi + L) + (1-x)\Pi \big) - 1 - \kappa(\lambda) \right], \quad (28)$$

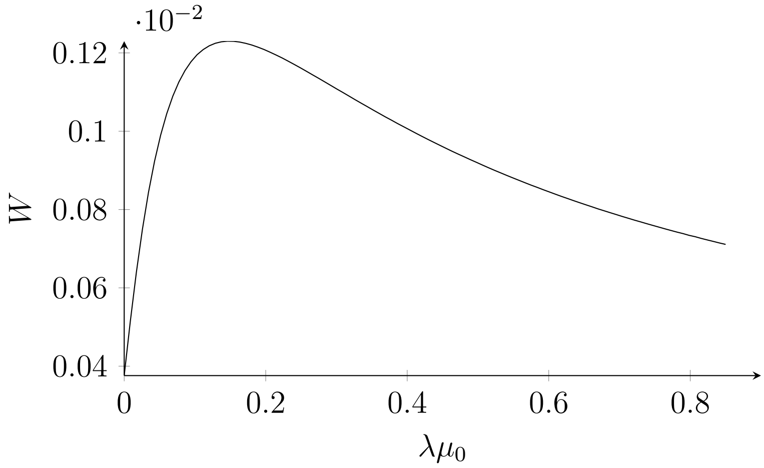
where $L \equiv \ell/(1-\beta)$ is the present value of a permanent dead-weight loss. Notice that in $e(\hat{s}, \gamma(\lambda))$ we are making explicit the dependence of γ on λ (see equation (15)). To explain this welfare expression consider Figure 5 when the latest technology available in the market attains a marginal cost z_1. The first term in square brackets captures the arrival of an obvious entrant, which occurs with probability $1-\alpha$ and who obtains a patent with probability λ. With probability $1-x$ the obvious entrant lands in a competitive niche, not affecting welfare. With probability x the obvious entrant lands in a monopolized niche and the incumbent takes it to court. The obvious entrant succeeds with probability μ_0, in which case, the market price goes down from z_0 to z_1. The area A depicts the profits of the replaced incumbent which are transferred to consumers as surplus. The area A brings no new social surplus. The area B, on the other hand, is the dead-weight loss associated with the patent protection given to the incumbent. With the arrival of the obvious innovation, this loss is transferred to consumers permanently, increasing welfare by L.

The second term in the expression represents the payoffs when the entrant is a genuine innovator and obtains a patent, which occurs with probability α. With probability x the entrant lands in a monopolized niche and gets challenged in court. The entrant wins in court with probability μ_1. In that case, the price goes down from z_0 to z_1. As before, the area A is transferred from the incumbent to consumers and the original dead-weight loss, B, is now captured by consumers permanently. In contrast, the entrant captures the area $C + D$ as profits. When a new innovation arrives in the future, these profits will be eventually transferred to consumers. That is, the welfare value created are the areas $B + C + D$ at perpetuity, or $\Pi + L$. Finally, when the genuine innovator lands in a competitive niche (that is, the existing price equals the marginal cost of the latest technology z_1), it does not get challenged in court and appropriates the area $C + D$ as profits. As before, these profits will eventually be transferred to consumers when a new innovation arrives. The entrant, thus, creates a welfare value Π. The area E is the new deadweight loss created, not appropriated by anyone until a new innovation successfully enters the market.

In order to understand better the welfare trade-offs in this model we can re-arrange (28) using the steady-state values of relevant objects provided in Lemma 1 we obtain

$$W = e \left[p(\Pi + L) - 1 - \kappa(\lambda) \right],$$

which is the expression presented in the main text. Figure 14 presents an example of an interior screening rate by the patent office when screening is free and the courts' behavior is exogenous.



Note: Parameter values are $\alpha = 0.1$, $\pi = 2.4$, $\ell = 0.13$ (i.e., $a = 24$ and $\delta = 0.9$), $\beta = 0.8$ and $\mu_1 = 0.85$.

Figure 14: Welfare with deadweight loss; interior maximum when screening is free.

The Judges' Problem We now analyze how a judge's endogenous decision changes when innovations are cost reducing and a dead-weight loss might arise. As explained in the main text a type I error arises whenever a genuine innovator is prevented from entering the market. In this scenario the type I error leads to a loss $E_I^{CS} = \Pi + L$.

The type II error represents the "loss" in social welfare when a firm with an obvious innovation is allowed to replace an active monopolist. This negative cost has now two components. First, there is a short run gain, derived from eliminating the deadweight loss that the incumbent generated. Second, there is the same dynamic effect explained in the benchmark case that increases the future value of the niche. That is,

$$E_{II}^{CS} = -L + \beta(w_M^{CS} - w_C^{CS}),$$

where the value of a monopolistic and competitive niche are respectively defined as

$$\begin{aligned} w_C^{CS} &= \beta w_C + e(\hat{s}, \gamma(\lambda)) \left[\alpha \left(\Pi + \beta(w_M - w_C)\right) - 1 - \kappa(\lambda)\right], \\ w_M^{CS} &= \beta w_M + e(\hat{s}, \gamma(\lambda)) \left[\alpha\mu_1(\hat{s})(\Pi + L) + (1-\alpha)\lambda\mu_0(\hat{s}) \left(L + \beta(w_C - w_M)\right) - 1 - \kappa(\lambda)\right]. \end{aligned} \tag{29}$$

The difference with the benchmark case is that, now, the value of a monopolistic niche depends on the dead-weight loss L. Each time a monopolist is replaced by another firm the deadweight loss associated to its innovation is eliminated and a discounted surplus L is transferred to consumers by means of a lower price. This effect increases w_M and narrows the difference between the value of a monopolistic and a competitive niche.

As a result, the introduction of a deadweight loss generates two opposing effects on the type II error. First, there is the static gain of eliminating the deadweight loss, enhancing the return from eliminating an existing monopolist; i.e., making the type II error more negative. The second effect, corresponds to the dynamic effect of reducing the incremental value of a competitive niche relative to a monopolized one (discussed in the previous paragraph). As the next lemma shows, the static effect dominates.

Lemma B.1. *For all $\hat{s}$ and γ, $E_{II}^{CS}(\hat{s},\gamma) < E_{II}(\hat{s},\gamma)$ where $E_{II}(\hat{s},\gamma)$ is given by* (14) *and*

$$E_{II}^{CS}(\hat{s},\gamma) = E_{II}(\hat{s},\gamma) - L\frac{(1-\gamma)[(1-\beta)+\alpha\beta(1-\mu_1(\hat{s}))e(\hat{s},\gamma)]}{(1-\beta)(1-\gamma)+\alpha\beta(1-\gamma+\gamma\mu_0(\hat{s}))e(\hat{s},\gamma)}.$$

Proof. Subtracting the value functions w_M and w_C in (29) and solving we obtain

$$\begin{aligned}\beta(w_M - w_C) &= \frac{[\beta(\alpha\mu_1(\hat{s})+(1-\alpha)\lambda\mu_0(\hat{s}))L - \alpha(1-\mu_1(\hat{s}))\Pi]e(\hat{s},\gamma)}{1-\beta+\beta(\alpha+(1-\alpha)\lambda\mu_0(\hat{s})e(\hat{s},\gamma)}\\ &= L\frac{\beta(\alpha\mu_1(\hat{s})+(1-\alpha)\lambda\mu_0(\hat{s}))e(\hat{s},\gamma)}{1-\beta+\beta(\alpha+(1-\alpha)\lambda\mu_0(\hat{s})e(\hat{s},\gamma)} + E_{II},\end{aligned}$$

where (14) was used in the last step. Replacing in $E_{II}^{CS} = -L + \beta(w_M - w_C)$ delivers the expression in the lemma. ■

We now formalize the details of the analysis presented in the main text for an individual judge's decision. Recall that when a judge chooses $s = 1$ no error is made, so the only cost is that related to her effort, $J^{CS}(1,\hat{s},\gamma) = c$. When a judge chooses effort $s = 0$, however, the cost is

$$J^{CS}(0,\hat{s},\gamma) = J(0,\hat{s},\gamma) + L(1-\gamma)\Delta(\hat{s},\gamma)/2,$$

where $J(0,\hat{s},\gamma)$ is defined in (17) and

$$\Delta(\hat{s},\gamma) = \frac{(1-2\gamma)(1-\beta)+\alpha\beta(1-\gamma)e(\hat{s},\gamma)}{(1-\gamma)(1-\beta)+\alpha\beta\left(1-\gamma+\gamma\mu_0(\hat{s})\right)e(\hat{s},\gamma)} \leq 1. \tag{30}$$

Proposition B.1 (Complementarity in the model with DWL)**.** *For a given $\hat{s}$, $\Delta(\hat{s},\gamma)$ is decreasing in γ. Consequently, $J^{CS}(0,\hat{s},\gamma)$ is decreasing in γ. That is, the presence of the term in L in J^{CS} reinforces the negative effect of γ on the best response function of an individual judge.*

Proof. From Proposition 2 we know that $J(0,\hat{s},\gamma)$ decreases in γ. We need to show that $\Delta(\hat{s},\gamma)$ decreases in γ. Start by substituting for $e(\hat{s},\gamma)$ in (30) to obtain

$$\Delta(\hat{s},\gamma) = \Phi(\hat{s},\gamma) - \Omega(\hat{s},\gamma), \text{ where } \quad \Omega(\hat{s},\gamma) = \frac{\gamma k_1}{(1-\gamma)[\Pi\alpha k_1 - (1-\mu_1(\hat{s}))(1-\gamma)]},$$

$\Phi(\hat{s},\gamma)$ is given by (18), and k_1 and k_2 are defined in the proof of Lemma 5. By Lemma 5, $\Phi(\hat{s},\gamma)$ decreases in γ. Hence, it is sufficient to show that $\Omega(\hat{s},\gamma)$ increases in γ. Differentiating

$$\frac{d\Omega}{d\gamma} = \frac{\Pi\alpha k_1^2 - (1-\mu_1(\hat{s}))(1-\gamma)\left(k_1+\gamma\mu_0(\hat{s})\right)}{(1-\gamma)^2\left[\Pi\alpha k_1 - (1-\mu_1(\hat{s})))(1-\gamma)\right]^2}.$$

The denominator is always positive. We use Assumption 1 (i.e., $\Pi \geq k_2/(\alpha k_1)$) to construct the following lower bound for the numerator

$$k_1k_2 - (1-\mu_1(\hat{s}))(1-\gamma)\left(k_1+\gamma\mu_0(\hat{s})\right) = \gamma\mu_0(\hat{s})(1-\gamma)(4-5\mu_0(\hat{s})) + (\gamma+\mu_0(\hat{s})-1)^2$$

where we have used $\mu_1(s) = 1-\mu_0(s)$. This expression is positive as $\mu_0(s) \leq 1/2$ for all s; proving that the derivative is positive. Therefore, $\Delta(\hat{s},\gamma)$ and, consequently, $J^{CS}(0,\hat{s},\gamma)$ decrease in γ. ■

B.2 Continuous Enforcement Effort

We previously assumed that a judge's effort, s, could take only two values. The judge could either exert no effort, 0, or full effort, 1. In this section we explore the case where effort is continuous; i.e., $s \in [0,1]$. As expected, the main results carry through.

To ease the exposition, we assume that the cost of exerting effort is quadratic. In particular $c(s) = \tilde{c}s^2/2$ where $\tilde{c} > 0$ is a scale parameter.[27] Building on Lemma 5, the objective function that a judge minimizes with respect $s \in [0,1]$ can be written as

$$J(s, \hat{s}, \gamma) = \Pi(1-\gamma)(1-s)\Phi(\hat{s}, \gamma)/2 + c(s), \tag{31}$$

where $\Phi(\hat{s}, \gamma) \in (0,1]$ is given by equation (18). This objective function is the continuous-effort analogous of (17). The next lemma characterizes an individualjudge's best response taking the possibility of corner solutions into account.

Lemma B.2. *For a given enforcement effort by other judges $\hat{s}$ and screening rate by the patent office, γ, a judge's best response is unique an given by*

$$s(\hat{s}, \gamma) = \min\left\{1, \Pi(1-\gamma)\Phi(\hat{s}, \gamma)/2\tilde{c}\right\}. \tag{32}$$

If $\tilde{c} > \Pi/2$, the best response is always interior; i.e., $s(\hat{s}, \gamma) \in [0,1)$. If $\tilde{c} \leq \Pi/2$ there exists values of $(\hat{s}, \gamma)$ for which the best response is $s(\hat{s}, \gamma) = 1$.

Proof. The proof that $s^*(\hat{s}, \gamma)$ is interior whenever $\Pi/2 < \tilde{c}$ was given in the main text. This solution is a minimum, as the second order condition is given by $\tilde{c} > 0$. For the case in which $\Pi/2 \geq \tilde{c}$, take $\gamma = 0$. Then, $(1-\gamma)\Phi(\hat{s}, \gamma) = 1$ and the first-order condition of the judge's problem (31) satisfies $\tilde{c}s - \Pi/2 < 0$ for all $s \in [0,1)$. Thus, $s^*(\hat{s}, 1) = 1$ is the unique solution to the minimization problem. ■

The objective function in (19) is strictly convex so when $s(\hat{s}, \gamma)$ is interior it characterizes the unique solution of the judge's problem. Additionally $s(\hat{s}, \gamma)$ is always positive so the only corner solution that may arise involves $s(\hat{s}, \gamma) = 1$. When the cost of effort is sufficiently high, $\tilde{c} > \Pi/2$, $s^*(\hat{s}, \gamma) < 1$ for all values of $(\hat{s}, \gamma)$. To see this observe that $\Pi/2\tilde{c} < 1$ and, by Lemma 5, $(1-\gamma)\Phi(\hat{s}, \gamma) \leq 1$.[28] When the cost of effort is not too high, $\tilde{c} < \Pi/2$, the corner solution $s(\hat{s}, \gamma) = 1$ arises for values of $(\hat{s}, \gamma)$ that make $(1-\gamma)\Phi(\hat{s}, \gamma)$ sufficiently close to one.

In the rest of this section we focus on the case $\tilde{c} > \Pi/2$ so that the solution is guaranteed to be interior. This assumption is analogous to Assumption 2 in the binary effort scenario.

Proposition B.2 (Screening complements enforcement)**.** *For a given screening rate by the patent office γ, there is a unique symmetric steady-state equilibrium in the enforcement game in which judges' effort is s^* such that $s(s^*, \gamma) = s^*$. In this equilibrium, an increase in the patent office's quality of screening increases judges' enforcement effort s^*. Thus, patent screening and patent enforcement are complementary.*

[27] Results can be easily generalized to an environment with a convex cost function (i.e., $c'(s) > 0, c''(s) > 0$ for $s > 0$) satisfying $c(0) = c'(0) = 0$.

[28] Under a general cost function $c(s)$, the sufficient condition for an interior solution is $c'(1) > \Pi/2$.

Proof. For a given γ, a symmetric interior equilibrium is given by

$$s^* = \Pi (1-\gamma) \Phi (\gamma, s^*) / 2\tilde{c}.$$

We prove that the symmetric equilibrium is unique. Define the function $F(s,\gamma) = \Pi (1-\gamma) \Phi (\gamma, s) / 2 - \tilde{c}s$. Every symmetric equilibrium is a solution to $F(s^*,\gamma) = 0$. To prove uniqueness, we show that dF/ds is negative; thus, for a given γ, F can only cross zero once. Differentiating

$$\frac{dF}{ds} = \frac{\Pi (1-\gamma)}{2} \frac{d\Phi}{ds} - \tilde{c} \leq \frac{\hat{\Pi}}{2} \left((1-\gamma) \frac{d\Phi}{ds} - 1 \right).$$

where $d\Phi/ds$ is given by (25). The inequality above follows from the assumption $\tilde{c} > \Pi/2$. We show that an upperbound of the parenthesis in the expression above is negative. The parenthesis is equal to

$$\frac{\hat{k}_1 \left[\hat{k}_1 \left(2\gamma (1-\gamma)^2 - \hat{k}_2^2 \right) + 2(1-\gamma) \left(\gamma (1-\gamma) k_3 + \hat{k}_2^2 (1-s) \right) \right] - \left(\hat{k}_2 (1-\gamma)(1-s) \right)^2}{\hat{k}_2^2 \left(\hat{k}_1 - (1-\gamma)(1-s) \right)^2}$$

where $k_3 = s + 2\gamma - 3 < 0$, and $\hat{k}_1$ and $\hat{k}_2$ are defined in the proof of Proposition 7. Because the denominator is positive, we need to show that an upperbound of the numerator is negative. The last term of the numerator is a subtracting a positive term. Thus, it is sufficient to show that the term in square brackets is non-positive. The first term of the square brackets, $2\gamma (1-\gamma)^2 - \hat{k}_2^2$, is negative. This can be readily verified using that $\hat{k}_2$ decreases in s. The second term of the square brackets is positive. Thus, the square brackets can be positive or negative. Using Assumption 1, we can use that the positive entry assumption implies $\hat{k}_1 > \hat{k}_2$, to find the following upper bound for the term in square brackets

$$\begin{aligned} &\hat{k}_1 \left(2\gamma (1-\gamma)^2 - \hat{k}_2^2 \right) + 2(1-\gamma) \left(\gamma (1-\gamma) k_3 + \hat{k}_2^2 (1-s) \right) < \\ &\quad - \left(2\gamma (1-\gamma)^3 (1-s) + (2 - \gamma (s+1))^2 ((1-s)\gamma + 2s(1-\gamma)) \right) < 0 \end{aligned}$$

Therefore, $dF/ds < 0$ and the symmetric equilibrium is unique.

To prove that the Patent office's effort is complementary to a Judge's effort we use the Implicit Function theorem; i.e.,

$$\frac{ds^*}{d\gamma} = -\frac{dF/d\gamma}{dF/ds}.$$

Observe that $dF/d\gamma = dJ/d\gamma$. From the proof of Proposition 2 we know $dJ/d\gamma < 0$. Consequently, we have that $ds^*/d\gamma < 0$, proving the result. ■

Proposition B.2 shows the robustness of the complementarity found in the case with binary enforcement effort. Better patent screening (a decrease in γ) increases the enforcement effort exerted by the judges. It is also interesting to observe that the multiplicity of equilibria is not present in this case. Specifically, with continuous enforcement efforts and after ruling out corner solutions with $s^* = 1$, we no longer have a coordination problem among judges.[29] Figure 8 depicts the equilibrium under various parameters and screening rates. The upper left panel shows the complementarity between patent screening and patent enforcement. The other two figures show how the equilibrium enforcement effort increases in Π and α.

[29]This suggests that the coordination problem in the binary efforts case is related to the manner in which the prospects of perfect enforcement by subsequent judges ($\hat{s} = 1$) fully removes the social value of type II error, reinforcing the incentive of an individual judge to choose $s = 1$. For $\hat{s} < 1$, type II error by an individual judge is still valuable at the margin.