COSTLY SCREENS AND PATENT EXAMINATION

Jonathan S. Masur

ABSTRACT

The United States Patent and Trademark Office has acquired a well-deserved reputation for inefficacy and inefficiency. Proposals for reforming the patent office have thus focused on improving the quality of patent review while decreasing its cost. Yet this view overlooks the valuable function performed by the high costs associated with obtaining a patent: these costs serve as an effective screen against low-value patents. Moreover, due to asymmetries in patent values, the costly screen is likely to select against socially harmful patents in disproportionate numbers. Although the patent office is the most prominent forum in which this type of costly screening operates, it is not the only one. In a variety of other contexts, the private costs of navigating an administrative process may complement the process itself in screening out unwanted participants.

1. INTRODUCTION

For decades, the United States Patent and Trademark Office (PTO) has received trenchant criticism regarding the manner in which it reviews patent applications and the costs it imposes upon applicants. Due in large part to the incentives the PTO places upon its own employees, patent office review has acquired a reputation as an extremely poor screen against non-novel or otherwise invalid patents (Jaffe & Lerner 2004; Kieff 2003; Lemley 2001; Merges 1999). The reasons for these failures have been equally well documented. Examiners spend on average only eighteen hours reviewing each patent, and their incentives are structured so as to bias them heavily in favor of granting patent applications (Lemley 2001, 2001).

1 Assistant Professor, University of Chicago Law School; jmasur@uchicago.edu. I thank Michael Abramowicz, Douglas Baird, John Bronsteen, Chris Buccafusco, Jake Gersen, Lee Fennell, Mark Lemley, Thomas Miles, Anne Joseph O’Connell, Mark Ramseyer, Pam Samuelson, Ted Sichelman, Matthew Stephenson, Lior Strahilevitz, Polk Wagner, and an anonymous referee for helpful comments. Jim Grant provided excellent research assistance. This work was supported by the David and Celia Hilliard Fund.
1500; Thomas 2001; Jaffe & Lerner 2004, 11–13). Worse still, the PTO’s inefficiency is accompanied by high procedural costs: an inventor will spend approximately $22,000 to obtain a patent.

Accordingly, arguments regarding the patent office have traditionally centered around whether patent examinations are worth the (considerable) expense; that is, whether the patent office should spend additional resources pursuing more rigorous examinations, or whether it should forgo these efforts entirely and scrutinize patents only after they have been granted and asserted against alleged infringers. In one respect, however, these discussions have been consistent: in all cases, they have weighed the costs of engaging in patent review against the substantive benefits of that review, as measured by the number of “bad” patents caught and rejected. According to this approach, the administrative costs of prosecuting a patent are simply the purchase price of the active scrutiny conducted by the PTO.

But this is not the only function performed by the patent system’s process costs. The high costs of prosecuting a patent force inventors to determine ex ante whether the property rights they might acquire are genuinely worth the expense. This ex ante private cost creates a type of costly screen: the patent applicant must decide whether the expected benefits of obtaining a patent, discounted to present value, exceed the costs of navigating the patent office process (Carroll 2004). This price barrier forces potential applicants to draw upon private information about the value of their inventions, information that the patent office is otherwise unable to obtain.

This article applies a costly screening model to the procedural operations of the Patent and Trademark Office.2 According to this model, patent examination is properly understood as a price-setting mechanism, not just as a referendum on the substantive merits of the patent.

The fact that patent office processes function as a costly screen is not, by itself, terribly interesting; it says nothing about what types of patents will be eliminated. And it is here that this article’s most significant contribution lies. Because of an asymmetry in how patents are valued, the PTO’s costly

---

2 This paper draws upon the standard costly screening models that have been profitably deployed in a variety of legal and economic contexts (see, e.g., Stephenson 2006; Gersbach 2004; Posner 2001; Stiglitz & Weiss 1989; Milgrom & Roberts 1986; Spence 1973). Within the patent context, a few prior articles have addressed the possibility that patents may be used by firms as an inexpensive means of signaling financial markets or other outsiders (Lemley 2000; Long 2002), but none develops a screening model of patent examination and explores which classes of patents will be selected against by such a screen.
screen will disproportionately select against patents that are harmful to overall social welfare, while leaving beneficial patents almost entirely untouched. The reason lies with the ways in which patents distribute wealth to their owners and to society at large, and with the strength and breadth of the patent right itself.

First, imagine dividing the universe of patents very roughly into “low” and “high” value types. High value patents represent significant, successful inventions; low value patents are commercially irrelevant. Next, consider the crucial distinction between the private value of a patented invention (what it is worth to the patent holder) and the public or social value of that invention (what it is worth to social welfare at large). Viewed across both dimensions simultaneously, patents fall into four conceivable categories.

First, there are high private value, high social value patents: these are valuable, novel inventions (new drug compounds, innovative computer circuits, etc.) that contribute something tangible to social well-being and might not exist but for the research incentives created by the patent system. They represent the paradigm case for patent rights. Second, there are patents with high private value and low or negative social value: these are minor or insignificant innovations that contribute little to public knowledge but lead to blocking patents and allow their owners to extract significant rents (Merges 1994). Third, there are patents of low private value and low (or negative) social value; this class of patents includes both discarded, unenforced patents that increase the search costs and risk imposed on commercial firms—the “patent thicket,” in popular parlance (Shapiro 2001)—and worthless, largely unenforceable patents usable only for extracting nuisance settlements (see Section 2.2.). And fourth, one could imagine patents of low private value and high social value.

The goal of the patent system is, of course, to weed out low or negative social value patents, increasing overall welfare by preventing them from seeing the light of day (or a courtroom). The PTO’s costly screen will not block high private/low social value patents because they are worth too

---

3 It is important to note that “low” need not mean “greater than or equal to zero.” Under certain circumstances, a patent can have negative value. This fact has great significance for the argument that follows. See Section 2 below.

4 Of course, patents are valued along a continuum, and it is a simplification to categorize them only as holding “high” or “low value. Nonetheless, this is a useful shortcut and one that later sections will describe in somewhat greater depth.
much to their putative owners. However, it will effectively select against low private/low (or negative) social value patents—one significant class of undesirable property rights. In this sense, substantive patent examination and the PTO’s costly screen serve as effective complements. The former may block many harmful high private/low social value patents that the costly screen does not reach; the latter selects against low private/low social value patents, where substantive examination is especially ineffective.

This benefit could be outweighed if the screen similarly selected against high social value patents, a welfare-enhancing category of inventions the patent system should be designed to promote. Yet here the costly screen exploits a crucial asymmetry between the private and public values of patented inventions: patents of low private value and high social value are almost entirely nonexistent. The monopoly rights conferred by patents ensure that any invention with high social value will also create significant private value for its inventor—that is, after all the purpose of granting patent rights in the first instance. Consequently, the costly screen will reduce or eliminate an important category of low and negative–social value patents without equally discriminating against a corresponding class of high social value patents. There is thus reason to believe that the patent office’s screen is welfare enhancing.

This is not to say, however, that the PTO’s costly screen represents a first-best solution to the problem of large numbers of low private value/low social value patents. The costly screen exists predominantly due to the attorneys’ fees required to prosecute a patent, and those fees are deadweight losses. It might be possible to improve on the current situation by substituting higher patent office fees for attorney-driven costs, if this could be achieved without further harming the already deficient patent examination system. Nonetheless, the status quo is likely superior to any reform that would reduce the cost of obtaining a patent without simultaneously imposing some other type of screen against harmful low private value/low social value property rights.

The costly screen imposed by the PTO’s process costs is particularly significant because of the confluence of two factors: the “substantive” patent examinations purchased by patent process costs are not terribly effective against low private/low social value patents, and the costly screen is able to exploit an asymmetry in patent valuations that permits it to select against predominantly welfare-diminishing patents. But the patent office is by no means necessarily the only setting in which these two phenomena might be present, and thus in which process costs might serve a similarly beneficial
function. More generally, the application of costly screening models to administrative procedure may reveal a wide variety of contexts in which ostensibly ineffective administrative processes actually serve as powerful costly screens. In legal settings ranging from environmental permitting, to immigration law, to landlord-tenant regulation, the private costs of navigating an administrative process may serve to eliminate many of the unworthy candidates that administrators themselves are unable to expose.

This novel approach to administrative process holds potentially great significance for questions of institutional design, at the patent office and elsewhere. Otherwise well-intentioned efforts to reduce the costs of obtaining a patent or streamline PTO operations may well have the perverse effect of permitting tens of thousands of worthless, damaging patents to see the light of day, to the detriment of inventors and consumers alike. And institutional design reforms in other administrative contexts aimed at limiting transaction costs or cutting apparently wasteful operations could unleash countless pernicious activities that a costly screen would otherwise block.

This article proceeds in four parts. Section 2 summarizes the operation of the Patent and Trademark Office, the pathologies surrounding patent examination, and the harmful social consequences produced by nuisance patents. Section 3 describes and analyzes how patent office procedures effectively impose a costly screening against low-value patents, despite the inadequacy of examination procedures themselves. Section 4 abstracts away from the particular context of the patent office and describes the operation of costly screens more generally and their function within three other important fields of administrative law. Section 5 concludes.

2. THE PATENT OFFICE: HISTORIC PROBLEMS AND CONVENTIONAL REMEDIES

The Patent and Trademark Office is plagued by agency problems related to the incentives it creates for its examiners. Patent examiners have little personal reason to resist the granting of invalid patents and significant private incentives to allow those patents to go forward. Examiners also spend very little time scrutinizing each patent. Consequently, patent attorneys have come to believe that they can effectively “wear down” even recalcitrant patent examiners with continuous appeals and refilings. These improperly granted patents can exact a social cost, dissuading firms from entering into markets or commercializing inventions and clogging the processes of innovation.
Suggested reforms to this system fall into two camps. Some scholars advocate investing greater amounts of money in more robust patent office review. Others, pointing to the high costs associated with patent examinations and the large percentage of patents that hold little or no commercial value, suggest scaling back (or even eliminating) the PTO examination process and moving toward a system of patent registration and strong ex post review in the courts or a redesigned administrative agency. Both groups, however, treat the expenses that the PTO and private parties must bear in prosecuting a patent solely as the cost of the active examination that takes place, to be avoided or minimized wherever possible. And there is reason to believe that the PTO’s examination procedures are not cost-benefit justified on those grounds.

2.1. Rational Examiners and Misguided Incentives

Stories of ridiculous, invalid, and obvious patents are legion (see, e.g., Merges 1999). In recent years the PTO has allowed patents on a stick, the process of toasting bread, and a method for swinging on a swing, to name just a few examples (Jaffe & Lerner 2004, 32–33). Far more importantly, however, the PTO has almost surely granted invalid patents on thousands, if not tens of thousands, of software, biotechnology, or integrated circuitry inventions (Jaffe & Lerner 2004, 37). These patents, on inventions that would have been obvious to scientists in the field or were anticipated by prior work, carry with them the potential to stifle innovation, to discourage firms from entering into useful markets, and generally to impede the optimal functioning of the American economy. Even if they are never litigated—indeed, especially if they are never litigated, and never see the inside of a courtroom—these “bad” patents impose significant deadweight losses and delays in precisely those industries in which rapid progression and the growth of small-scale market participants are most important.

These patents exist first and foremost, of course, because of the rents that they permit their owners to charge. But they are allowed to exist also because of the inadequacies and pathologies of the procedures employed by the patent office to screen them out. The patent office describes itself as existing to provide a service to patent applicants, who are its “customers” (U.S. Patent and Trademark Office 1994), and states quite plainly that its mission is “to help our customers get patents” and “to ensure strong intellectual property for all Americans” (U.S. Patent
and Trademark Office 1994, 1997)—hardly a celebration of the office’s role as examiner. In any other federal agency this might be mere rhetoric, a paean to the idea of a friendly, facilitative government bureaucracy. At the patent office it is an insight on the institutional role exemplified in the procedures that the office has created to process applications and the incentives placed upon the key actors within the system, the patent examiners.

Each patent application filed with the PTO is referred to a single patent examiner who holds plenary authority over the application for nearly all of its life (Merges & Duffy 2007, 48–54). After she has examined the patent, the examiner must choose whether to grant or reject the patent application. As an initial matter, rejecting a patent application is more difficult and time-consuming for the examiner than granting one. If the examiner grants the application, little further procedure is necessary—the examiner simply announces that she is allowing the application to mature into a patent. If the examiner rejects the patent, however, she must provide a statement of the reasons for her rejection, identify the relevant prior documents or inventions (the PTO refers to these as “prior art references”) and the section of the Patent Act that has caused her to reject the application, and generally explain the rationale behind her actions.5

Patent examiners receive salary bonuses based on the number of patent applications that they are able to process (Jaffe & Lerner 2004, 116). This fact, by itself, might skew the examiner’s incentives, inclining her toward acceptance rather than rejection based upon the differing workloads and time expenditures required. But the problem is in fact far greater and lies with how the PTO understands what it means to fully process an application.

Unlike a patent grant, an examiner’s decision to reject a patent application does not end the matter. First, the patent examiner cannot issue a “final” rejection on the first go-around.6 If the examiner initially rejects the patent, the applicant is entitled to request a re-examination in front of the same examiner.7 These preliminary rejections are known as “office actions,”8 and they consist principally of correspondence from the examiner to the applicant explaining which claims cannot be granted and for

5 37 CFR § 1.104(a)(2) (2007) (“The reasons for any adverse action or any objection or requirement will be stated in an Office action….”).
what reasons. At this point, the patent applicant may choose to abandon the application, though in practice few do. Instead, most applicants elect to respond to the patent examiner’s concerns, revise the application (often editing or redacting certain claims), and request re-examination.9 The patent examiner is then again faced with the choice of whether to accept the application or reject it.

After this second examination, the examiner may choose to issue a “final” rejection of the application, though she need not do so.10 In theory, the examiner and the applicant could engage in an infinitely iterated series of preliminary rejections and re-examinations, and indeed many patents are the subject of three or four office actions before they are finally accepted or rejected.11 Yet even if the examiner issues a final rejection of an application, the matter is not closed. If the applicant does not wish to abandon the invention, she may file a continuation application, which is little more than a request for re-examination (presumably involving edited claims or new arguments) attached to an additional filing fee.12 The patent application remains before the same examiner as if the “final rejection” had not been genuinely effective. Moreover, there is no limit to the number of continuation applications that an applicant may file.13 If the applicant is

---

9 “Request” is a bit of a misnomer; the examiner has no choice but to consider the application a second time.

10 37 CFR § 1.113(a) (2007) (“On the second or any subsequent examination or consideration by the examiner the rejection or other action may be made final…”) (emphasis added).

11 This estimate is based upon conversations with patent prosecutors at a number of law firms, principally Kirkland & Ellis LLP and Schiff Harden LLP. Notes on file with author.

12 35 U.S.C. § 120 (2007). The applicant can also appeal the decision directly to the Board of Patent Appeals and Interferences (“BPAI”), which can overturn the examiner’s decision and send the patent back to the examiner for further consideration. 35 U.S.C. § 134 (2007). If the applicant loses before the BPAI, she then holds the right to appeal the decision to the Federal Circuit. 35 U.S.C. § 141 (2007). The applicant may also bring a civil action in federal district court against the director of the patent office, seeking essentially the same relief, though few choose this route. Id. § 145.

willing to pay the necessary costs, the examiner has no way of rejecting the application decisively.

Meanwhile, as these various appeals are taking place, the application resides in the examiner’s file as an open matter, rather than a processed one, and the examiner does not accrue credit toward her next bonus (Jaffe & Lerner 2004, 136). Worse, if her decision is overturned she faces the prospect of expending even more time on an application that has not provided her with any meaningful return. The rational, self-interested examiner thus has a tremendous incentive to grant the vast majority of patent applications—a self-interest that is only buttressed by the organizational tenets of the patent office itself. Not surprisingly, scholars estimate that patent examiners spend, on average, only eighteen hours scrutinizing each application (Lemley 2001, 1500; Thomas 2001, 310).

A number of separate studies have attempted to determine the proportion of patents approved by the PTO, with varying results. One early study found that the PTO grants approximately 97 percent of the patents it examines (Quillen & Webster 2001), a figure that was later revised downward to 85 percent (Quillen, Webster, & Eichmann 2002); another study found a grant rate of approximately 75 percent (Lemley & Sampat 2008). Irrespective of the precise figure, there is ample evidence to indicate that patent examiners are performing poorly when it comes to weeding out invalid patent applications (Merges 1999; Schaafsma 2004).

14 One interesting and suggestive study found that patent approval rates spike in September—the month in which the PTO’s accounting year closes and examiners are awarded bonuses for processed applications. See Gajan Retnasaba (2008).

15 This is in contrast to other administrative organizations, such as prosecutors’ offices, that structure internal cultures and incentives so as to mitigate the effect of rational self-interest. Prosecutors themselves have an incentive to dismiss cases or settle them quickly, for short sentences, in order to dispose of their workload and maximize leisure time. Prosecutors’ offices counter this incentive by creating cultures that value longer sentences and higher conviction rates and evaluating line prosecutors on those grounds. See generally Stephanos Bibas (2004).

16 Confusion may be due in part to difficulties with the data that the PTO itself provides. The PTO website reports that between 1963 and 2005, the patent office received 4,016,707 new utility patent applications and issued 3,891,905 patents, which would be the equivalent of an astounding 96.8 percent rate of success. See U.S. Patent Statistics Chart, Calendar Years 1963–2006, available at http://www.uspto.gov/go/taf/us_stat.htm. However, the number of actual patent applications filed must be well in excess of the 4,016,707 figure once continuations and other spin-off applications are included. The Lemley & Sampat (2008) study, which avoids relying on the PTO’s self-reported data, is likely closer to the actual result.
Of course, the fact that PTO examiners likely allow too many patents does not mean that they entirely neglect their obligation to scrutinize applications. Even patents that are improvidently granted may be substantially narrowed in the course of examination, as examiners cancel or restrict the least tenable claims. Yet this narrowing has differential effects on various classes of patents in various circumstances. In some cases, the examiners’ narrowing function may significantly mitigate the systemic costs created by improperly granted patents. In others it may have little impact. The section that follows describes some of these costs imposed by “bad” patents, and Section 3 returns to the question of which types of patent-related costs will be most affected by examiners’ efforts at narrowing overly broad patents.

2.2. The Costs of “Bad” Patents

Patents provide inventors with limited monopoly rights largely in order to incentivize innovation (Chisum et al. 1998, 6; Merges & Duffy 2007, 253–256). This tradeoff is accepted with respect to valid patents on novel inventions, but even invalid, improperly granted patents on pre-existing technologies (that involved no useful innovation) have the power to dissuade potential competitors from entering a market and stunt investment in further research (Ayres & Klempere 1999, 1018–1020; Heller & Eisenberg 1998). This is the case regardless of whether or not the invalid patents are ever asserted; the threat posed by the existence of those patents is enough to raise barriers to market entry (Leslie 2006).

Invalid patents augment the costs to prospective market participants in three ways. First, a market entrant must investigate the intellectual property that exists in the field and make some preliminary inquiry as to those patents’ validity. This investigation, even if cursory, can be quite expensive. Second, invalid patents can hamper a firm’s ability to raise capital (Federal Trade Commission 2003) or write contracts with potential customers (Leslie 2006, 125–127). Financial markets will be wary of firms that may not be sustainable because they traffic in infringing products. Customers will hesitate before forming business relationships that may expose them to suits for contributory infringement (Borkin 1950, 641) and resist relying upon suppliers who may be shut down or driven out of the market by a lawsuit. Finally, firms will have reason to fear the cost of defending a lawsuit.

for patent infringement, not to mention the threat of having to pay licensing fees or royalty damages. Patent lawsuits of any length impose asymmetric costs upon the participants: it is easier and less costly for patent holders to prove infringement than it is for alleged infringers to prove invalidity (Jaffe & Lerner 2004, 152), largely because patents arrive in court accompanied by a legal presumption that they are valid. Litigation, even relatively nonmeritorious litigation, thus presents a substantial threat.

A single, significant patent of plausible validity can cause these types of problems for a nascent competitor. Importantly, though, a large quantity of frivolous, obviously invalid patents within the field can create the same sorts of barriers to entry (Leslie 2006, 132–137; Ayres & Parchomovsky 2007, 6–17). As an initial matter, search and information costs for the entering firm will be high regardless of whether these patents are ever enforced, as the market entrant is forced to comb through a dense “patent thicket” in order to ascertain the boundaries of existing property rights (Ayres & Parchomovsky 2007, 6–17; Merges & Duffy 2007, 615–616). It may also be difficult and costly for new firms to credibly signal necessary third parties such as banks, investors, and customers that a set of threatening patents are invalid, particularly when those third parties are not experts in the relevant technologies.

Most significantly, nascent market participants might face higher up-front costs if litigation uncertainties and information asymmetries force the firm to pay small licensing fees or settlements to a series of patent-holders who choose to file nuisance lawsuits (Leslie 2006, 133). Firms that face the prospect of being nickel-and-dimed by the owners of multitudinous dubious patents may well choose to refrain from investing in the development of new technologies in the first instance (Meurer 2003, 515). In addition,

---

18 35 U.S.C. § 282 (2007). However, the chorus in favor of altering this legal rule is growing. See Lichtman & Lemley (2007).
19 See also Bresnick v. U.S. Vitamin Corp., 139 F.2d 239, 242 (2d Cir. 1943) (Hand, J.) (describing a patent as a “scarecrow” that can deter competition by its very existence); but see Brunswick Corp. v. Riegel Textile Corp., 752 F.2d 261, 265 (Posner, J.) (“[A] patent known to the trade to be invalid will not discourage competitors from making the patented product or using the patented process, and so will not confer monopoly power….”). Judge Posner may be correct that a patent must be of at least “colorable” validity in order for it to be used as a means of exerting monopoly power, but see Leslie (2006, 133), but his analysis does not speak to the possibility that the asymmetric transaction costs involved in patent litigation will enable the holder of a plainly invalid patent to extract small payouts from market entrants. I return to this point in greater detail in Section 3.2.3.
the nuisance lawsuits themselves can produce significant deadweight losses; litigants expend thousands of dollars in transaction costs to prosecute and settle nuisance lawsuits worth $20,000 or less. I return to this point in greater detail in Section 3.2.3. For the moment, suffice it to say that even plainly invalid patents can impose significant social costs through sheer force of numbers.

2.3. Traditional Reforms

In response to the inadequacies of the patent office and the costs of bad patents, scholars have advanced a number of proposals designed to shore up that failing agency and provide a more effective screen against non-novel and potentially harmful patents. Some scholars have recommended increasing PTO funding in order to enable the office to hire more examiners and spend a greater amount of time on each patent (e.g., Lemley, Lichtman, & Sampat 2005, 12–13; Allison & Hunter 2006; Jaffe & Lerner 2004, 203; Ghosh & Kesan 2004). Another, smaller cadre has asserted that patent examinations should be eliminated altogether, with the patent system reverting to a simple system of registration akin to the copyright regime (e.g., Mossoff 2007; Kieff 2003).

These assessments of the patent system share a common feature: they treat the cost of obtaining a patent and the quality of the patent examination as necessary tradeoffs—the one exists only to create the other. Accordingly, some scholars have advocated that patent office fees be kept as low as possible in order not to impede applicants from filing (e.g. Dreyfuss 2006, 1577). None of these approaches considers the possibility that the high cost of prosecuting a patent might itself have a beneficial effect on the

20 See also Patent Reform Act of 2009, S. 515, 111th Cong. (2009). Many of these proposals are coupled with suggestions for meaningful inter partes post-grant administrative review, mechanisms by which potential infringers can challenge a patent’s validity without undertaking expensive litigation in federal courts (Farrell & Merges 2004; Jaffe & Lerner 2004; Lemley 2001). Some even recommend a multi-tiered system of patent review in which applicants can opt for one of several levels of PTO scrutiny with correspondingly strong ex post presumptions of validity (Lemley, Lichtman, & Sampat 2005, 12–13; Osenga 2005). And even more exotic proposals abound, including suggestions for tradable patent rights that will limit the number of patents in force at any given time by compelling patentees to bid on a finite pool of litigation rights (Ayres & Parchomovsky 2007, 22–39). But see Abramowicz (2007) arguing that government is ill-suited to determining when patent auctions should be held. These proposals for inter partes or multi-tiered review are in many cases quite compelling, and the theory set forth here can serve a complementary role to any or all of them.
quality of patents issued by screening out some significant number of socially harmful property rights. That is not surprising; no prior account develops a model of which sorts of patents a costly screen might deter. The next section aims to supply that model.

3. PATENT PROCEDURES AS COSTLY SCREENS

The administrative expense involved in obtaining a patent functions as a costly screen against low-value property rights. The screen will deter prospective applicants from filing for patents when they believe that property rights in their inventions will be worth little: in the tens of thousands of dollars or less. Of course, applicants will be concerned only with the private value of their inventions—what the patents will be worth to the applicants themselves. But the costly screen exploits an asymmetry within the distribution of value across patents. Patents of low private value will predominantly offer only low (or negative) social value as well. It is thus likely that the PTO’s costly screen enhances social welfare by selecting disproportionately against socially harmful property rights.

3.1. The Costs of Obtaining a Patent

Patents are relatively expensive to obtain. An initial patent application on a relatively complex technology—a semiconductor or biotechnology patent, for instance—will typically cost between $11,000 and $15,000 when prepared by a reputable law firm (Kasper 2008). Each preliminary rejection by the PTO generates an “office action” to which the patentee must respond, to the tune of approximately $4,000 in additional attorneys’ fees per office action (Kasper 2008, 7; Macedo 1990). Filing a continuation patent after a “final” rejection by the PTO is even more expensive and can cost as much as $10,000 in attorneys’ fees alone (Kasper 2008, 7; Macedo 1990). Once patent office fees and other attorneys’ costs are figured into

21 These figures and those that follow were confirmed in a number of independent conversations with attorneys at a variety of law firms, principally Kirkland & Ellis LLP and Schiff Harden LLP. Notes on file with author.

22 The Patent Office charges a variety of small fees for prosecuting a patent. See, e.g., C.F.R. § 1.16(a)(1), (k), (o) (filing fees); 35 U.S.C. § 41(a)(2) (same); 37 C.F.R. § 1.18(a) (issuance fees); 35 U.S.C. § 41(a)(1)(B) (same); 35 C.F.R. § 1.16(h), (i) (fees for claims); 37 C.F.R. § 1.16(j) (same).
the equation, an average patentee will spend approximately $22,000 to successfully prosecute a patent application.\textsuperscript{23}

Importantly, however, these costs are not consistent across all types of patents. Rather, costs will scale based on the extent to which the patent borders on invalidity and, even for valid patents, the crowdedness of the technological field from which the invention derives. This cost scaling is not due to any deliberate action by the PTO, but instead is the result of an interaction between the costs involved in responding to initial rejections by a patent examiner and the informational forces that drive rejection. Most obviously, patents of suspect validity are more likely to garner repeated office actions from the PTO as the examiner questions the invention’s patentability. Similarly, the more heavily congested a technical field is with prior inventions (particularly patented inventions\textsuperscript{24}), the more likely the examiner will find art that calls one or more of the claims into question.\textsuperscript{25} Consequently, transaction costs will be significantly higher for inventors who attempt to push through questionable patents, or who attempt to patent inventions in heavily commercialized fields in which those patents might do the most harm. In addition, repeated office actions will delay a patent’s issuance, eating into the twenty-year patent term that begins to run on the day a patent application is filed.\textsuperscript{26}

In effect, then, the very administrative processes that allow patentees to “wear down” examiners serve to increase the barriers to entry for the least desirable patentees. If the patent system is crudely successful at screening for invalid or damaging patents, it is not only (or primarily) because examiners are actually denying those patents. Rather, the procedural mechanisms that exist in the name of “customer service” exert a bias against the filing of applications on unpatentable inventions in the first instance.

\textsuperscript{23} This figure is based on calculations undertaken by the author based on a set of representative patents. Notes are on file with the author and available upon request. In 2001, Mark Lemley estimated the average cost at $10,000 to $30,000 (Lemley 2001, 1498). If anything, then, the estimates here may be overly conservative.

\textsuperscript{24} Examiners have better access to patents than they do to prior art in any other form (Jaffe & Lerner 2004, 145–149). They may in fact be unaware of substantial quantities of important prior art that has not been reduced to patent form.

\textsuperscript{25} The semiconductor and computer fields are typically understood to be heavily patented, while the pharmaceutical industry is generally thought to involve fewer overlapping property rights (Jaffe & Lerner 2004; Lemley 2007).

3.2. Asymmetries in Private and Social Values

By itself, however, the fact that patent office-generated administrative costs will act as a costly screen says little about what sorts of patents will be screened out and whether the screen is, by any measure, normatively desirable. After all, if the screen is not deterring harmful patents, it exists purely as a senseless source of transaction costs. I argue here that there is good reason to believe that the PTO’s screen produces meaningful welfare benefits.

The analysis proceeds in two steps. First, consider the universe of conceivable patents, by which I mean all patents that currently exist and those that inventors could conceivably file. Divide this universe into two categories: “low” and “high” value patents. These categories are defined by the cost (approximately $22,000) of prosecuting a patent at the PTO. Accordingly, the PTO’s costly screen will likely block low value patents, but it will not deter firms from filing for high value patents. These categories are necessarily quite rough: a low value patent is one with value of the same order of magnitude as the cost of obtaining a patent—in the low tens of thousands of dollars or less (including patents of negative value). A high value patent is one whose value exceeds that threshold substantially.

Next, consider the distinction between the private value of a patented invention (what it is worth to the patent holder) and the public or social value of that invention (what it is worth to social welfare at large). There are four possible “flavors” of patent when viewed across both dimensions simultaneously. First, there are high private value, high social value patents; these are the valuable, novel inventions (new drug compounds, innovative computer circuits, etc.) that contribute something tangible to social well-being and might not exist but for the research incentives created by patents. They represent the paradigm case for the patent system.

---

27 This is not to say that firms will always be able to determine precisely the expected value of their inventions ex ante. There will undoubtedly be some amount of uncertainty and error in these calculations. However, it seems reasonable to assume that sophisticated firms that are repeat players in an industry will have some reasonable estimate as to the likely value of their inventions. After all, the analysis here relies only on estimates as to the order of magnitude of the patent’s value.

Studies of the large number of low-value patents are not to the contrary (Moore 2005). Firms are calculating the ex ante expected value of patents. If there is any uncertainty in this calculation, firms will end up filing for some patents that they believed would be valuable but which did not pan out. In addition, individuals within firms may be excessively optimistic about their inventions (Armor & Taylor 2002, 334).
Second, there are patents with high private value and low or negative social value; these are minor or insignificant innovations that contribute little to public knowledge but lead to blocking patents and allow their owners to extract significant rents (Merges 1994). Third, there are patents of low private value and low or negative social value; these are quite common and come in a variety of shapes and forms, which I discuss in greater detail below. And fourth, one could imagine patents of low private value and high social value. Table 1 illustrates these four potential types of patents graphically.

Table 1. Four Possible Patent Types

<table>
<thead>
<tr>
<th>1. High private value/High social value</th>
<th>2. High private value/Low social value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Low private value/High social value</td>
<td>4. Low private value/Low social value</td>
</tr>
</tbody>
</table>

When a potential applicant considers whether to file for a patent, she will be concerned only with the expected private value of the patent. Accordingly, the PTO’s costly screen will only select against the low private value patents in categories 3 and 4. Yet here the costly screen exploits a significant asymmetry within the distribution of patents across these categories. Patents of low private value and high social value—the fourth category—are extremely rare (or even nonexistent). The monopoly rights that patents confer ensure that almost any invention with high social value will also create significant private value for its inventor. Consequently, the low private value patents selected against by the PTO’s costly screen will necessarily hold only low—or, more importantly, negative—social value as well.

One final clarifying note is in order. Because the categories of high and low private value and high and low social value discussed below are defined by the cost of obtaining a patent, the argument here is one of categorization: there are significant numbers of patents that fall into three of those categories, but not the fourth. The sections that follow describe the types of patents properly classified under each heading.

3.2.1 High Private Value, High Social Value Patents

Patents exist in order to encourage research and innovation—the purpose of granting a patent right is to permit an inventor to capture a proportion of the commercial value of her invention, and thereby to encourage those
inventions in the first instance. The paradigmatic patent, then, is one that is both valuable to the private holder (high private value) and covers an invention that is valuable to the public at large (high social value). These types of patents come in many forms—patents on useful new drug compounds, patents on innovative semiconductor devices, etc.—but they will share three common characteristics. First they must be at least plausibly valid, and thus plausibly enforceable as property rights; and second, they must claim inventions (or important components or subparts of inventions) that are commercially viable and useful in a market economy. A patent that satisfies those two conditions is privately valuable—its owner will be able to extract rents either through licensing or through production of the patented good. If the patent is to have social value—if the invention behind it is to be social welfare–enhancing—a third condition must be satisfied: the patent must describe inventions that are genuinely new and thus contribute some socially valuable knowledge that did not previously exist.

The patent system is designed to promote precisely this type of high private/high social value patent. And while the PTO’s costly screen will make these patents slightly more costly to obtain, it will likely block few or none of them. Twenty-two thousand dollars is a meaningful amount of money, but it represents little more than a rounding error in comparison to a truly valuable intellectual property right. Any marketable new product, or any important component or improvement related to a pre-existing...
product, will undoubtedly have a market value well beyond $22,000. The cost of obtaining a patent is unlikely to discourage researchers who believe that their work will lead to useful, marketable inventions.

### 3.2.2 High Private Value, Low Social Value Patents

The question of whether a patent has high private value for its owner and the question of whether the availability of a patent has spurred socially productive research and innovation are not always coterminous. Even where a patent does not involve novel, socially productive research and development, it may nevertheless be privately valuable in a number of ways if it is plausibly valid and commercially relevant. The patent might be deployed offensively, with the intention of collecting awards for infringement or licensing fees (Moore 2005, 1522–1524; Bar-Gill & Parchomovsky 2003, 1867); it might hold value as defensive mechanisms for protecting commercial products from competition or from suit for infringement (Barton 2002; Shapiro 2001, 121); or it might be valuable as a signal to deter potential competitors (Long 2002, 651–653; Lemley 2000, 144). The patent literature is rife with examples of patent plaintiffs who succeeded in collecting substantial infringement judgments based on patents that were never commercialized or even publicized, and which were not based on any genuine innovation. As long as the patent can be plausibly asserted against other firms doing business in the marketplace, it will be privately valuable to its owner.

Yet if the creation of the patent involved no socially beneficial research, its existence will prove socially detrimental in the net. These types of patents raise transaction costs and business risks for commercial firms that must negotiate with patent holders, defend against infringement claims, and run the risk of being litigated out of business. In the absence of socially productive research and development, these patents provide no corresponding social benefit to offset the transaction costs and hindrances to competition they create. They have high value only for their owners, and negative value for society at large.

Although it would be socially beneficial to eliminate these types of high private/low social value patents, the PTO’s costly screen will not serve as a meaningful barrier against them. The cost of obtaining a patent is a small

---

32 These types of plaintiffs are colloquially known as patent trolls, on the theory that they collect tolls for crossing bridges that they did not build themselves (Allison, Lemley, & Walker, 2009, 14).
fraction of the value to be realized from a property right of this type. Even a single successful lawsuit based on a plausibly valid but uncommercialized patent can net millions or hundreds of millions of dollars. Patents that appear to have value of this type will not be blocked by a costly screen set several orders of magnitude lower.

Yet despite the fact that the costly screen has little effect against these types of patents, substantive patent examination can nevertheless impact this class of patents substantially. This is the case even if the PTO will only occasionally reject them outright (as is likely the case). Even if a patent is improperly granted it might still be substantially narrowed during the examination process. This narrowing can greatly diminish the force of what might otherwise be a high private/low social value patent. Patents possess high value only if they can be read to cover commercially successful products. Any narrowing of scope that diminishes a patent’s commercial reach (or calls that patent’s validity into question) will curtail the patent’s usefulness in litigation or licensing and reduce the number and size of the awards that the patent holder can collect. In effect, then, patent examination may succeed in converting some number of putative high private/low social value patents into low private/low social value patents by stripping the patents of much of their scope and force. It is in this sense that a costly screen can never fully substitute for substantive patent

33 For example, in 2006 Blackberry maker Research in Motion agreed to pay NTP $612.5 million to settle an infringement claim despite the fact the US Patent & Trademark Office had notified both parties that the patents in question would likely be rejected after a final review (Kelley 2006). In 2007, Apple settled a patent lawsuit with Burst for $10 million (after Microsoft had settled a similar suit for $60 million) despite the fact that the patent at suit was extremely broad and possibly obvious (Lee 2007).

34 For instance, Lucent Technologies won a verdict against Microsoft in the tens or hundreds of millions of dollars based on a patent that covered “An arrangement for use in a computer… comprising means for displaying… a pattern of information fields and for identifying for each field a kind of information to be inserted therein….” U.S. Patent No. 4,763,356. A jury found that the “calendar” function on Microsoft Outlook infringed this patent, despite the fact that the patent appeared directed at a different sort of technology. Lucent Technologies, Inc. v. Gateway, Inc., 580 F.3d 1301 (Fed. Cir. 2009).

Similarly, in 2005 a firm called Pinpoint, Inc. sued Amazon.com based on a patent entitled a “System Method for Scheduling Broadcast of and Access to Video Programs and Other Data Using Customer Profiles.” U.S. Patent No. 5,758,257. Pinpoint claimed that Amazon’s customer rating software (the programs that determined what customers “like you” have purchased) infringed this patent, despite the fact that the patent was clearly directed at software for scheduling television broadcasts. The district court eventually rejected Pinpoint’s claims, but only after another ruling favorable to Pinpoint had been reversed on a technicality. Pinpoint, Inc. v. Amazon.com, Inc., 369 F. Supp. 2d 995 (N.D. Ill. 2005).
examination. However, a screen can serve as a useful complement to a system of substantive examination, as the next sections demonstrate.

3.2.3 Low Private Value, Low Social Value Patents

(a) The Patent Thicket. Unlike the patents discussed above, there is an entire class of low private/low (or negative) social value patents that the Patent Office’s costly screen will select against. These patents come in a variety of forms, but two flavors predominate. The first are those patents that comprise the “patent thicket” described in Section 2.2: those essentially worthless patents that are allowed to lie fallow and are rarely enforced, but that nonetheless drive up search costs and increase litigation risk for firms seeking to do business in the relevant market (see Section 2.2). The patent thicket is one species of anticommons, the well-theorized environment in which excessive numbers of overlapping property rights increase the transaction costs of doing business (Heller 1998). A number of scholars have previously noted the operation of an anticommons in patent law and the costs that over-patenting impose on commercial firms (Shapiro 2001; Heller & Eisenberg 1998). The patents that form the thicket have very low value to their owners—they are valuable only to the extent that their owners wish to keep competitors out of the marketplace. Accordingly, they almost certainly diminish social welfare by retarding competition without producing any meaningful inventive quid pro quo.

It is difficult to accurately measure the social costs created by the patent thicket because they are typically internal to the firms that incur them and hidden from public view. In addition, many of these costs may come in the form of forgone market opportunities or research avenues, and these speculative costs are of course highly uncertain. Nonetheless, the estimates that exist place the social costs of the patent thicket in the hundreds of millions of dollars (Lemley 2001; see also Leslie 2009; Leslie 2006, 115), and those figures are likely conservative. Of course, these are the costs created by the existing patent thicket. Understanding the effect of the PTO’s costly screen requires investigating the counterfactual: how much greater would the social costs of the patent thicket be if obtaining a patent was effectively free? This adds yet another layer of uncertainty to the inquiry, and this paper makes no claim to being able to provide a full answer. Nonetheless,

35 These types of measurement problems are endemic to patent law, where information on costs and benefits is difficult to obtain. Accordingly, the figures reported here should be viewed only as suggestive of the overall condition of the patent environment.
without any sort of costly screen the number of granted patents would likely increase substantially, and with them the search and uncertainty costs imposed upon firms seeking to do business.

(b) Nuisance Patents. The second major flavor of low private/low social value patents is the class of patents that are useful principally as mechanisms for filing nuisance lawsuits. Several scholars have identified patent law as an area ripe for exploitation by nuisance lawsuits (Sudarshan 2009; Moore 2007; Lemley 2001). The reason lies with the informational asymmetries inherent to patent litigation and the manner in which the costs of litigation are distributed.

Some patent lawsuits involve two competitors within an industry, parties that have likely eyed one another warily for some time and kept close watch on each other’s patent portfolio. But a substantial percentage of these actions are initiated by a solo inventor or patent holding company with no commercial ventures beyond the exploitation of its intellectual property portfolio (Allison, Lemley, & Walker 2009). At the inception of such an action, plaintiffs—particularly non-commercial plaintiffs—enjoy a substantial informational advantage over their targets. Plaintiffs know the content of their own patents, as well as other information relevant to the patents’ validity, such as prosecution histories. The defendant’s allegedly infringing device is an actual physical product that exists in the world whose relevant characteristics may be easily ascertainable. By contrast, the infringer very likely knows nothing of the patent and its claims (much less its prosecution history), and may have little information regarding the relevant prior art that preceded the patent.

In order to cure this informational asymmetry, most targets of an infringement suit will immediately commission an opinion letter from outside counsel to determine whether the patent is valid and the firm’s device infringes it.36 The purpose of this letter is two-fold. The letter is meant both to inform the potential infringer of the strength of the patent-holder’s case and to guard the potential infringer against later claims of willful infringement37 by supplying the basis for a good-faith

---

36 This is standard practice within the field (see Lemley 2001), a fact that I confirmed repeatedly in the course of interviews and conversations with patent attorneys at several firms.

37 The patent statute allows courts to assess treble damage penalties against willful infringers. 35 U.S.C. § 284 (2008) (“the court may increase the damages up to three times the amount found or assessed.”); In re Seagate Tech., LLC, 497 F.3d 1360, 1371 (Fed. Cir. 2007) (setting forth the modern standard for determining when infringement has been willful).
belief that the patent is not infringed. According to the defendant, even a patent defendant who is quite familiar with the plaintiff’s patent may feel it necessary to immediately obtain an opinion letter.)

As with all legal work relevant to patent law, these opinion letters can be quite expensive. Major law firms typically charge at least $8,000 to $12,000 to write opinion letters covering technologically sophisticated patents and inventions, and those costs can reach $30,000 or more if the technologies involved are sufficiently complex or the patents and products sufficiently numerous (Kasper 2008; Lemley 2001). For potential defendants, then, every colorable assertion of infringement carries with it a nuisance value in the neighborhood of $10,000: this is the amount that the infringer will have to spend at the outset in order simply to understand the contours of the putative case against it. Even after surmounting this hurdle, accused infringers must confront the asymmetries that make patent lawsuits more expensive to defend than to bring (Lichtman & Lemley 2007). Commercial firms—particularly smaller firms that do not possess the resources to defensively litigate test cases—thus run a meaningful risk of becoming targets for nuisance lawsuits (Moore 2007, 90–91; Bone 1997).

Much like the patent thicket, these types of nuisance lawsuits can impose significant costs on commercial firms. As I described in Section 2.2, threats of multiple small lawsuits can dissuade firms from entering new markets (Meurer 2003) and increase the costs of capital (Leslie 2006; Ayres &

---

38 See, e.g., Delta-X Corp. v. Baker Hughes Production Tools, Inc., 984 F.2d 410, 414 (Fed. Cir. 1993) (opinion letter provides near-impenetrable defense to charges of willful infringement); Nickson Industries v. Rol Manufacturing Co., 847 F.2d 795, 800 (Fed. Cir. 1988) (same). An accused infringer has no affirmative duty to seek an opinion letter if it wishes to avoid liability for willful infringement, In re Seagate Technology, LLC, 497 F.3d 1360, 1371 (2007), but the chances of a finding of willful infringement increase dramatically when an infringer has not obtained an opinion letter, and so nearly any colorable accusation will trigger a request for the opinion of counsel.

39 These estimates were similarly confirmed in conversations with attorneys at a number of law firms, principally Kirkland & Ellis LLP and Schiff Harden LLP.

40 This is not to say that potential defendants would always pay $10,000 to make every patent lawsuit disappear. Targets for nuisance lawsuits have incentives to send credible signals to potential future accusers that they will not be easy marks by litigating claims aggressively rather than settling them. See, e.g., Coffee (1986, 712–713); cf. Fearon (1994); Schelling (1956, 283–284) (“Concession not only may be construed as capitalism, it may mark a prior commitment as a fraud, and make the adversary skeptical of any new pretense at commitment.”). A small number of companies behave in this fashion; Wal-Mart is one known example. See, e.g., Keeton v. Wal-Mart Stores, Inc., 21 F. Supp. 2d 653, 660 & n. 13 (E.D. Tex. 1998).
Parchomovsky 2007). Each forgone potential market opportunity creates costs for consumers who must pay higher prices or are deprived of some good. Nuisance lawsuits also impose transaction costs as firms expend resources in filing and settling them, even where they do not proceed to trial (Leslie 2006). For commercial firms, particularly the smaller firms that are especially vulnerable to harassing litigation, nuisance lawsuits can generate substantial business expenses.

Not surprisingly, the net social welfare costs attributable to these types of nuisance lawsuits are difficult to measure. Many of the costs stem from forgone competition and other hidden business activities. In addition, most estimates group the costs from nuisance lawsuits with the costs imposed by more substantial lawsuits and licensing deals (e.g. Lemley 2001; Bessen & Meurer 2005; Allison et al. 2004), and so reliable measures of the costs they create are difficult to obtain. Nonetheless, it seems reasonable to believe that nuisance lawsuits (and the threat of nuisance lawsuits) impose substantial social costs, perhaps even in the hundreds of millions of dollars yearly (Leslie 2006; Lemley 2001; see also Sudarshan 2009; Bessen & Meurer 2005; Allison et al. 2004), despite the imprecision of these estimates. Again, these are estimates of the costs imposed by existing nuisance lawsuits, and the value of the PTO’s costly screen must be judged by comparison to the hypothetical universe of costless patenting. Nonetheless, the available evidence seems to indicate that the costs will be substantial.

It is worth noting that these nuisance lawsuits—and the patents behind them—are quite distinct from the high private/low social value patents described in the previous section. There will certainly be many socially worthless patents that are plausibly valid and sufficiently commercially important that they can be used to extract significant settlements or licensing fees, often measuring in the hundreds of millions of dollars for a single patent. The costly screen will have a negligible effect on the rates at which they are applied for and granted. Yet at the same time there exists a thriving market for genuinely nuisance-value patents: patents of such dubious validity that they can only be used to extract minor nuisance settlements priced below the cost of performing even a cursory evaluation of the patent. (The presence of nuisance lawsuits in patent law mirrors its operation in a variety of other

41 All of these costs are described in greater detail in Section 2.2, above.
42 For examples, see sources cited in note 33 above.
legal contexts (Rosenberg & Shavell 1985; Bebchuk 1988), including securities litigation (Alexander 1991).) Again, the point is a definitional one. Nuisance-value patents (as defined here) exist, and they exist in numbers likely large enough to generate hundreds of millions of dollars in costs for commercial firms (Meurer 2003; Leslie 2006).

(c) Low-Value Patents in Combination. The patents that comprise the “thicket” and those that give rise to nuisance lawsuits represent intellectual property at its very worst, deterring firms from entering markets or developing new products and consuming litigation resources while incentivizing essentially zero productive innovation. These patents are little more than carriers for transaction costs. They are, however, likely blocked by the PTO’s costly screen in substantial numbers. The upfront costs of obtaining a patent forces firms and inventors to at least consider whether an application is worth filing before adding another useless patent to the thicket. And when patents cost more to obtain than they can be used to extract in one or two nuisance settlements, they become substantially less attractive as a business tool and less open to exploitation (Bone 1997).43

Moreover, the costly screen is even costlier, and thus more effective, against these types of patents. Many of the more insidious patents described here hold only low private value because they are not plausibly valid.44 And for a patent to impose social costs it must bear some relevance (or resemblance) to an active commercial field. The class of low private/low social value patents is thus composed predominantly of dubiously valid, commercially relevant property rights—precisely the flavor of patents that will encounter the greatest number of hurdles during PTO examination. (This phenomenon is described in greater detail in Section

---

43 This is not to say that nuisance lawsuits will never be profitable, or that firms will never pursue questionable patents with the intent only to extract such settlements. A firm may be able to garner more than one quick payout with each patent, though at the same time it will not necessarily be capable of coercing targets—especially repeat players—into paying even inexpensive blackmail. Because of the costs of obtaining a patent, a firm cannot count on being able to turn a profit, or even recoup its investment, by threatening some number of small, meritless suits; it must actually believe that it has an invention worth commercializing or a valid patent in a commercially useful field before a patent application becomes worth the cost of prosecution.

44 Some patents will hold small private values because they are commercially insignificant—the patent on a method for swinging on a swing, for instance (Jaffe & Lerner 2004, 32)—but these patents are typically irrelevant from a social welfare perspective as well.
Accordingly, the costs of examination will be highest for these patents, and the costly screen will select against them in greater numbers. Importantly, the PTO’s substantive patent examination will be of little use in curbing these types of patents. Consider the narrowing function performed by patent office examination. Where the substantive scope of a patent is important—as it is for high private/low social value patents—this narrowing will affect the patent’s value substantially. But examiners’ work to narrow the reach of the patents they grant will have a much more modest impact on the social costs created by low private/low social value patents. When it comes to this category of patents, substantive scope is essentially irrelevant—by hypothesis, the patent cannot withstand even limited scrutiny. Rather, these patents create social costs simply by their very existence: both the patent thicket and the threat of nuisance lawsuits depend entirely upon large volumes of largely inapplicable patents with only a passing resemblance to the commercial products they affect. This is not to say that patent examination is worthless; the PTO may reject some fraction of low private/low social value patents, and it may narrow others to such a degree that they no longer appear even vaguely commercially relevant. But it will not exert the same force as it does against patents whose value depends on their commercial reach.

The costly screen thus serves as an important complement to substantive patent office review. The latter primarily targets high private/low social value patents; the former will eliminate a substantial number of low private/low social value patents. Of course, the Patent Office’s costly screen will hardly bar all of these low-private value, low-social value patents. Tens of thousands of such applications are filed yearly, and many of them are

45 Owners of many of these low-private value, low-social value inventions will opt for trade secret protection as an alternative to the overly costly patent system. Such a substitution should be no less beneficial to social welfare than if the inventions were simply eliminated. Low-private value, low-social value patents impose costs because of their existence as property rights, without which there can be no “thicket” to raise information costs and no basis for nuisance lawsuits. The shift to trade secrets eliminates these costs entirely. To be certain, trade secrets can impose their own set of costs: the loss of information that would otherwise be publicly disseminated, and the expensive steps that firms might take to protect them by restricting access both internally and externally (Lemley 2008). Yet these costs are likely to be negligible here. By hypothesis, the “trade secrets” behind low-private value, low-public value inventions are either not particularly secret, or not particularly related to trade. (That is to say, they are either not novel ideas or not commercially viable.) In either case, there is little to be lost if they are hidden from the public, and no firm will expend particular effort in maintaining them as private information. For this class of intellectual property, the “intellectual” aspects are essentially irrelevant; it is the status as property right (or not) that matters.
granted eventually (Lemley 2001, 1528). Yet without a costly screen—if, for instance, the PTO were to move to a registration system (see Mossoff 2007; Kieff 2003), or if the patent office were to reduce its fees to the minimum possible (Dreyfuss 2006, 1577)—the problem would likely be far worse. By selecting against this class of patents, the process costs perform a beneficial function, one that may eliminate greater numbers of these harmful patents than the substantive examination that the process costs are themselves used to purchase.

3.2.4 Low Private Value, High Social Value Patents

The benefits provided by the PTO’s costly screen would be quickly counterbalanced if the screen similarly selected against low private/high social value patents—patents that were worth little to their owners but contributed socially productive research and innovation. But patents are not symmetrically distributed across the four categories of value.

With few exceptions, low private/high social value patents do not exist. Any truly novel, commercially relevant invention—i.e., any socially productive invention—will give rise to a privately valuable patent on that invention. The patent system is designed to accomplish precisely this end: patents allow inventors to capture a substantial portion of the wealth created by their inventions. The Supreme Court’s extension of the scope of patentable subject matter to cover “anything under the sun made by man”46 only accentuates this fact.

Even inventions that might appear at first glance to fall into this category are not truly low private value/high social value in the sense meant here. Consider, for instance, “orphan” drugs—pharmaceutical inventions for which patent protection has expired (or nearly so) (Sichelman 2010, 386–387). Orphan drugs are surely low private value/high social value inventions in the most literal sense: these drugs would be valuable to society if manufactured and distributed, but no firm can make a great enough profit from them to render their development commercially worthwhile (Roin 2009). Yet this fact is not at all attributable to the cost of obtaining a patent on

46 Diamond v. Chakrabarty, 447 U.S. 303, 309 (1980) (internal citation and quotation marks omitted) (permitting patenting of newly made life forms); see also Diamond v. Diehr, 450 U.S. 175 (1981) (permitting patenting of mathematical and computer algorithms). The patentability of business methods is currently pending before the Supreme Court, see Bilski v. Doll, 129 S. Ct. 2735 (2009), but for the moment the PTO has been permitting patenting of business methods so long as they claim a general-use computer. See, e.g., Ex Parte Dickerson, 2009 WL 2007184 (B.P.A.I. Jul 9, 2009). If this rule were upheld, it would allow for patents on an extremely broad range of inventions.
the drugs. These drugs are unprofitable because the costs of obtaining FDA approval are so high—in the tens or hundreds of millions of dollars. In addition, once one pharmaceutical company has obtained FDA approval, others will be able to free-ride off of that approval without undergoing the same expense and compete away the first company’s profits (Roin 2009). The $22,000 cost of obtaining a patent is simply irrelevant to the calculation.

There will, of course, be minor exceptions to this rule—the transformative idea that does not directly give rise to an “invention,”47 or the peculiar patent that creates wealth that cannot be captured commercially. But these patents will be the rare outliers. Unlike the other three categories of private/public value relationships, there is no true class of low private/high social value patents.48 The asymmetry may not be absolute, but it is pronounced.

It is also possible that patents function in some cases as lottery tickets: an inventor might file for large numbers of patents, hoping (but not knowing) whether one will become valuable (Scherer 2001, 11). Ex ante, each individual patent might therefore be worth little to the inventor. Even at first glance, however, this theory does not seem to do justice to inventors and patent holders, at least on the valuation scales relevant here. These conceptions of large quantities of uncertainly valued patents credit inventors with little ability to discern the worth of innovation in their own commercial fields, contrary to evidence that patent applicants have “a fairly good sense ex ante as to which of their patents will be the most valuable (Burk & Lemley 2009, 52).” And here the inventor’s valuation need not be terribly fine-grained; the only salient question is whether the patent is worth only tens of thousands of dollars or substantially more.

Perhaps more importantly, even if the notion of patents as lottery tickets is an accurate representation of reality, it is not clear that it is one best left in place. Massive quantities of low-value patents impose significant


48 This will be the case even if one subscribes to the “portfolio” theory of patents (Parchomovsky & Wagner 2005). According to the portfolio theory, patents in many industries are more valuable in groups-bundled into “portfolios” than singly. This theory is meant to describe the patenting behavior of major firms with multimillion- or billion-dollar research and development budgets. The portfolio theory would indicate the existence of low private value/high social value patents only if there existed an invention that was worth a relatively small amount of money (for instance, $1 million) but required 50 patents to protect it. This is farfetched, and it is not the type of phenomenon that Parchomovsky and Wagner’s sophisticated theory was meant to predict.
negative externalities upon other firms seeking to do business in the same markets (Ayres & Klemperer 1999; Heller & Eisenberg 1998; Leslie 2006; Ayres & Parchomovsky 2007, 6–17; Merges & Duffy 2007, 615–16). The imposition of a costly screen may be forcing inventors to invest additional resources in acquiring information about the putative value of their inventions and cause them to be more selective in choosing which to file.49

Table 2 summarizes the relationships between private and social value for various types of patents. As the top row indicates, patents that carry high private value—and will be therefore worth obtaining despite the costly screen imposed by PTO procedures—can come in a variety of forms, only some of which are socially valuable. An expensive screen set in the tens of thousands of dollars will not select against socially beneficial or socially deleterious patents, so long as those patents carry significant private worth. At the same time, however, there are essentially no patents of low private value and high social value for the costly screen to bar; any invention involving a serious technological breakthrough or the creation of a commercially viable product or process will necessarily grant its holder a valuable monopoly right.50 Only patents of low private value and low or negative social value—precisely those patents most likely to diminish social welfare—will be meaningfully affected by the cost of PTO procedures.

It remains impossible to know whether the process costs involved with patent examination are justified in the aggregate. After all, every applicant—including those with valid patents and useful inventions—is forced to expend significant resources to obtain a patent. Nonetheless, there is good reason to believe that the costly screen imposed by the PTO’s process costs, coupled with the substantive examination purchased

49 This will prove impossible only when patent filing patent must necessarily precede systematic investigation of the invention’s commercial worth, most notably (and perhaps only) as with patents on new pharmaceutical compounds, which are filed before FDA trials on those drugs begin. See generally Roin (2009). There, whatever effect the PTO’s costly screen may be having, it is far from debilitating; the pharmaceutical industry is “often described as the patent system’s greatest success story.” Id. at 1; see also Eisenberg (2005) (“Patent law traditionally takes the lion’s share of credit for motivating investments in drug development.”).

50 It is possible that a badly drafted patent application on a significant technology will result in a low-value property right being conferred upon the inventor. For instance, an inventor might develop a useful technology but draft his patent application in such a way that it is easily evaded by competitors. While these sorts of weak patents would undoubtedly hold only modest private value, a costly screen will not deter inventors from seeking them. At the time of filing, the inventor does not realize that his patent is weak and thus will believe that it carries greater social value than it actually does.
by those costs, serves as a better filter against social welfare–diminishing patents than the PTO’s flawed examination would alone.

### 3.3. Tradeoffs and Second-Best Solutions

This article has attempted to describe the manner in which the patent office’s examination costs function as a costly screen against low private value patents. The article has argued further that these procedures may be normatively desirable from a social welfare perspective, in that the costs of obtaining patents are exceeded by the benefits of preventing low private/low social value patents from proliferating. Yet it is almost certain that this system of process cost-based screens is not a first-best solution to the problem of welfare-diminishing patents.\(^{51}\) It is a curious feature of the current patent system that the preponderance of the costs imposed against applicants are levied in the form of fees paid to third-party attorneys.\(^{52}\) In the alternative, the PTO could require applicants to pay substantially heightened fees to have a patent examined and granted. Unlike the procedural costs of shepherding an application through the patent office, patent fees are not deadweight losses; the PTO could simply bestow them upon future inventors in the form of research grants, in a tax-and-transfer system designed to properly align parties’ incentives.

<table>
<thead>
<tr>
<th>High private value</th>
<th>Low private value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High social value</strong></td>
<td><strong>Low or negative social value</strong></td>
</tr>
<tr>
<td>Commercial products; improvements; major components</td>
<td>Blocking patents; valid patents involving little novel research</td>
</tr>
<tr>
<td><em>extremely rare</em></td>
<td>Nuisance patents; minor inventions</td>
</tr>
</tbody>
</table>

\(^{51}\) For a seminal analysis of second-best solutions, see Lipsey & Lancaster (1956).

\(^{52}\) The explanation for this conceivably inefficient structure may lie in the political economy of the patent system. The patent bar is the largest cohesive political actor with a vested stake in the patent process, and the patent bar can be expected to oppose any change in PTO procedures that diminishes the role of patent attorneys. The problem is exacerbated by the specialization within the field: patent prosecutors, who represent applicants before the PTO (see http://en.wikipedia.org/wiki/Patent_prosecutor#United_States) do not typically represent clients in subsequent patent litigation. The patent bar will thus tend to oppose reforms that moderate the role of attorneys before the PTO even if they are traded off against increased post hoc litigation in federal court.
A number of scholars have quite rightly advocated increasing the fees charged to patent applicants (Lemley, Lichtman, & Sampat 2005; Osenga 2005; Jaffe & Lerner 2004).\textsuperscript{53} But these scholars have understood increased fees as a means of purchasing greater scrutiny for patents, not as a complement to such scrutiny. This has obscured the possibility that the costs of obtaining a patent may be doing as much work to prevent the issuance of socially harmful patents as the substantive examination itself.\textsuperscript{54} Other scholars have suggested moving to a system of registration akin to copyright, under which applicants would pay only very modest fees to obtain a patent (Mossoff 2007; Kieff 2003). These proposals focus on the transaction cost savings that would be realized from eliminating substantive patent examination, but they ignore the likely harmful consequences of simultaneously erasing the PTO’s costly screen. Because the costly screen targets social costs against which substantive patent examination is particularly ineffective, reforms that would decrease PTO fees to the lowest possible level (e.g. Dreyfuss 2006) seem inadvisable.

As legislators and administrators lever up or down the quantity of patent procedures in the course of one or another reform, they would be well advised to understand that they are simultaneously adjusting the costs imposed upon applicants and thus, crucially, the incentives those applicants face with respect to patents of questionable validity and value. Even if the optimal level of patent examination is zero, the optimal ex ante financial barrier to patenting likely remains much higher. Substantive examination and the costly screen serve as complements to one another: the former can narrow (and thus defang) potential high private/low social value property rights, while the latter selects against low private/low social value patents. The first-best solution, then, is likely some combination of examination and screen, with both generated to the greatest possible degree by higher patent fees, rather than transaction costs paid to patent attorneys. Until Congress sees fit to reorient the patent system along

\textsuperscript{53} Other commentators have suggested heightened ex post renewal fees as a means of thinning the patent thicket (see, e.g., Ayres & Parchomovsky 2007, 18–22), but these increased fees would impact only truly abandoned inventions and have no measurable effect on patents destined for use in nuisance lawsuits.

\textsuperscript{54} In addition, as transaction costs increase, the examination system could threaten to eliminate even high social value patents. The question of where best to set the costly screen involves consideration of the possibility that too high a barrier will select against some socially valuable inventions, as well as the notion that too low a screen will permit too many socially harmful property rights.
such lines, the PTO’s administrative procedures—and the costs they impose—will continue to perform a useful screening function.

4. ADMINISTRATIVE COSTS AS INFORMATION-FORCING BARRIERS

As the preceding Sections have argued, the costly screen imposed by the PTO’s process costs is particularly significant because of a confluence of two unlikely factors. First, the “active” examinations purchased by these process costs are substantively ineffective, which is to say that they generate high error costs when attempting to weed out invalid patents. And second, the costly screen is able to exploit an asymmetry in patent valuations that permits it to select against predominantly welfare-diminishing patents while leaving the majority of welfare-enhancing patents in place. This Section suggests that this theory of process costs may be more generally applicable across a variety of administrative contexts. The same essential dynamic may operate in several other fields: due process protections for employees subject only to “for-cause” termination and summary-process evictions; the obtaining of pollution permits; and numerous types of immigration visas, as well as citizenship status and even residence within the United States. These cases are not as severe; the administrative processes involved may be more efficacious than patent examination. But in all cases, the screen-creating costs of navigating the administrative system appear to complement and augment the screening value of the procedures themselves.

4.1. Due Process and Summary Process

4.1.1 Employee Termination Hearings

Employees may possess the right not to be fired except “for good cause” either as a matter of contract or, in the case of some federal, state, and municipal employees, as a matter of law. Before an employer may

55 The PTO has only limited authority to set its own fees, and any major adjustment to the fee schedule requires action by Congress (Rai 2009, 2067).
56 See, e.g., 55 ILCS 5/3-7012 (2007) (“Except as is otherwise provided in this Division, no deputy sheriff in the County Police Department, no full-time deputy sheriff not employed as a county police officer or county corrections officer and no employee in the County Department of Corrections shall be removed, demoted or suspended except for cause, upon written charges filed with the Board by the Sheriff and a hearing before the Board thereon upon not less than 10 days’ notice at a place to be designated by the chairman thereof.”).
discharge an employee subject to these protections, the employer must provide the employee with a hearing before a neutral arbiter and demonstrate that good cause for termination exists.57 In some cases, the employer may also be barred from depriving the employee of a salary before the hearing has concluded.58

Such hearings are not necessarily walkovers for employers. An employee may obtain representation, muster effective witnesses and evidence, and present a strong case that her behavior and performance were within the firm’s or the state’s acceptable boundaries. But employers have a set of systemic advantages stemming from their comparative size and the fact that they are repeat players within the system (cf. Calabresi & Cooper 1996; Galanter 1974).59 Employers understand what level of proof is necessary for success in this type of case, having brought many such actions. Employers are familiar with the limited cast of arbiters who will make the decisions. And employers have the financial capacity to hire better attorneys, where necessary. As a result, employee due process hearings are likely to be biased to some extent in the employer’s favor, and thus substantively ineffective to some meaningful degree.

With these advantages, however, come a number of asymmetric costs. In most cases, the employer must create and fund the hearing board, paying the hearing officers’ salaries60 and providing all of the other accoutrements that attend what is in essence a full-blown courtroom hearing.61 Furthermore, in some limited cases the employer must pay the employee’s wages during the pendency of the hearing (or the suspension that precedes it).62 Under most contractual or legislative arrangements these wages are legally recoverable if the employer succeeds in

59 In some instances, employees will be represented by collective bargaining units who are also repeat players, which may serve to mitigate these advantages to some extent.
60 See, e.g., 55 ILCS 5/3-7003 (2007) (stating that each Illinois county is responsible for paying the salary and expenses of the members of the boards established to conduct due process hearings for county employees).
61 The hearings in these cases are not minor affairs. They are conducted on the record, involve paper filings and live testimony, and frequently result in written decisions.
62 See Gilbert, 520 U.S. at 931; Loudermill, 470 U.S. at 545.
terminating the employee, but in practice the employee may be judgment-proof. Finally, as the difficulty of a case increases, or as the employer’s position becomes less certain, the employer may have to opt for higher-skilled—and thus more expensive—attorneys.

Consequently, scholars have suggested that these expensive due process protections may not be worth the cost to the employer or the employee. The employee will undoubtedly have to bear some of the cost of her due process rights in the form of reduced wages or other benefits, and the hearings may not be as valuable as the employee might hope because of the employer’s inherent tactical advantages. Better, perhaps, for both parties to eliminate the procedural rights and split the savings between them.

This narrow focus on the results of the pre-termination process and its administrative cost ignores the screening function that this cost performs. Due process costs force an employer to assess whether the harm that the employee is causing to the enterprise exceeds the transactional expense of terminating her. Irrespective of what “good cause” actually means or what a hearing board may decide, the administrative process sets a misfeasance threshold for the employment contract: an organization will move to discharge an employee only when the employee’s actions threaten substantial harm to the organization—or when an alternative employee would supply a substantially greater benefit—to a degree that well exceeds the administrative costs of termination. This is private information that only the organization—not the employee, and certainly not the hearing officer—possesses, and a costly process of removal forces the employer to disclose the information in the service of its own self-screening.

63 Cf. Goldberg v. Kelly, 397 U.S. 254, 264 (1970) ("[T]ermination of aid pending resolution of a controversy over eligibility may deprive an eligible recipient of the very means by which to live while he waits.").

64 See, e.g., Ellis v. Sheahan, 412 F.3d 754, 758 (7th Cir. 2005) (Posner, J.) (suggesting such an arrangement).

65 Just as some firms will litigate nuisance suits, despite the fact that litigation costs exceed settlement costs, in order to send signals of intransigence, it is certainly possible that some employers will pursue disciplinary actions against employees whose minor acts of misbehavior don’t cross this threshold in order to deter further such actions. But these punitive measures will likely be the exception, rather than the rule, given the other means of promoting productive behavior available to employers.
Moreover, this barrier is self-enforcing and essentially costless. When an employer decides not to take disciplinary action against an employee, it need not initiate an administrative proceeding, it need not hire an attorney, and it need not compensate the members of the administrative board that would hear the case. Like the cost of filing for a patent, it is the latent threat of having to pay for a hearing that forces the private party to screen ex ante.

If the termination of an employee of long standing imposes costs on society that the employer is not forced to bear, this costly screen may be welfare-enhancing as well. As with patents, there are probably few employee terminations that are worth little to the employer but a great deal to society at large—such a situation is difficult to imagine. More likely, any highly socially beneficial firing will be privately beneficial as well; the terrible employee who threatens general harm poses an even more substantial risk to the business or organization. As the organization’s interest in terminating the employee (the private value) shrinks, so too will the public good (social value) created by allowing the firing.

4.1.2. Housing Evictions

A similar dynamic operates in the context of housing evictions. Before evicting a tenant who has breached a lease, a landlord must summon the tenant into court and prevail before a neutral magistrate. The landlord may not simply cease performing her end of the housing contract by locking the tenant out (see Dukeminier et al. 2006, 382–408). This process is meant to be “summary,” and thus less costly for both landlords and tenants, but it is nevertheless characterized by many of the asymmetries of employment due process hearings: it is generally more costly for landlords than for tenants (though landlords are not asked to fund the courts); and

---

66 Which is to say that the salaries paid to hearing officers will reflect this diminished workload.

67 This is not to say that these sorts of due process rights are necessarily cost-benefit justified, or that employers and employees would not do better to bargain them away or move entirely to a system of at-will employment. Compare Epstein (1984) with Summers (1967). I take no position on these larger questions. The point is merely that the procedural costs associated with a form of administration may function more effectively as a screening device—and thus constitute a more valuable protection for employees—than the administrative process itself.

landlords, as repeat players, have systematic advantages when cases are contested and heard.

Laws that force landlords into court have been criticized in recent years, both as wealth transfers from law-abiding tenants to delinquent tenants and as stigmatizing devices that permanently taint tenants who acquire court records that future landlords can discover (Strahilevitz 2008). These criticisms are surely valid to some degree, and prohibitions on landlord self-help may do more harm than good to tenants or to the public as a whole. But the value of the administrative process to tenants lies not only in the results that tenants are able to obtain, but in the costs it imposes upon landlords as well. Landlords must decide ex ante whether removing a troublesome tenant is genuinely worth the expense. Tenants who do not reach this threshold—and are thus selected out by the landlord’s costly screen—will neither have a black mark placed upon their records nor create additional costs to be borne by landlords or other tenants. Again, if evictions involve significant negative externalities, and if those externalities increase as the value to the landlord of eviction decreases, such a screen may do significantly more good than harm.

4.2. Environmental Permits

Before a firm may construct a new source of pollution, such as a factory, it must navigate a dizzying array of federal and state environmental laws, install pollution-controlling technologies, and obtain a wide variety of permits from state and, in some cases federal, regulators. In order to comply


70 Again, I take no position on the underlying substantive questions.

71 A full analysis of this question is well beyond the scope of this paper, but there is a plausible story to be told that eviction valuations function in precisely this way. Any high social value eviction—for instance, the eviction of a disruptive tenant who is violent and engaged in illegal activity—likely holds high private value as well; the disruption strikes most heavily at other nearby tenants who are often under the auspices of the same landlord. The paradigmatic low private value/low social value eviction might be a tenant who is not disruptive but is delinquent on rent; turning such a tenant out onto the street might lead to crime and social disruption. This is the category of eviction against which a costly screen will select.

The danger posed by such a screening mechanism is that tenants will exploit the procedural costs involved with eviction by breaching their leases in minor ways, up to the point of making eviction worthwhile. Like any transaction cost, then, the costly screen could inhibit efficiency-enhancing transactions by enabling unnecessary bad behavior. Nonetheless, this danger may be less pronounced in residential housing than in other contexts. Tenants are likely to be highly risk averse—the downside risk of miscalculating and being evicted is substantial, and renters are often people with little margin to spare—and thus less inclined to push their luck.
with the mandates of the Clean Air Act, for instance, firms may be required to install cutting-edge technology and conduct continuous air quality monitoring for a period of one year in order to “determine the effect which emissions from any such facility may have ... on air quality.” The Clean Water Act imposes an entirely separate set of mandates, and other federal and state regulatory bodies may place further demands on prospective polluters.

The pollution-controlling devices that firms must install are certainly expensive, but they are far from the only source of cost involved. In addition, the administrative procedures themselves—the process of obtaining permits, filing monitoring reports, and ensuring compliance with overlapping regulatory regimes—can be extremely costly for polluting firms. For instance, the construction of a new chemical plant, involving the emission of scores of different chemicals from a variety of different points, can give rise to “stunningly complex” regulatory requirements and engender legal fees that run to the millions of dollars, even if the plant’s construction is never challenged in court (Campbell-Mohn, Breen, & Futrell 1997, 818).

The primary purpose of these administrative procedures is, of course, to ensure that concentrations of air and water pollutants remain at acceptable levels. But the high cost of compliance with environmental laws can serve a secondary purpose by weeding out those polluting activities that may not be cost-benefit justified, or at least may stray close to the borderline.  


73 Various provisions of the Clean Air Act require polluters to install technology that achieves the “lowest achievable emission rate” (42 U.S.C. § 7503(a)(2) (2007)), or represents the “best available control technology” (42 U.S.C. § 7479(3) (2007)). Compliance certification can be no small matter. The relevant technology standards can be rather amorphous, and compliance often relies on the subjective judgment of state regulators. See, e.g., id. (“The term ‘best available technology means an emission limitation based on the maximum degree of reduction of each pollutant ... which the [state] permitting authority, on a case-by-case basis, taking into account energy, environmental, and economic impacts and other costs, determines is achievable for such facility....”); see also Campbell-Mohn, Breen, & Futrell (1997, 820–22).


75 Id. § 7475(a)(7) (2007); see also Campbell-Mohn, Breen, & Futrell (1997, 820).


77 I do not mean to suggest here that environmental permitting processes are ineffectual in the sense of being easily evaded, but rather that they will not always be well suited to measuring the relevant quantity, which for present purposes I take to be overall social welfare.
Consider the potential construction of a factory. Suppose that the factory will cost $c$ to construct (exclusive of any costs related to controlling pollution) and produce benefits $b$. $C$ and $b$ are private information known only to the firm; the regulating agency cannot easily discover how much profit some new project is expected to produce. In the presence of environmental regulations, the firm must undertake compliance actions (installing scrubbers, obtaining permits, etc.) at a cost of $a$ and produces pollution that imposes a social cost of $d$. If $b > c + a$, the firm will choose to construct the factory; if $b < c + a$, it will not. But the factory is only justified in terms of overall welfare if $b > c + a + d$; society must bear the remaining pollution-related externalities, but the firm need not. Imagine a situation in which $b$ and $c$ are very similar—in other words, the factory has positive but small value net the costs of construction. If $d > b - c$ (or, to rearrange, if $b < d + c$)—in other words, if the social cost of pollution exceeds the private benefit from constructing the factory—these factories will do more overall harm than good and should not be built. If $a$ is small, they may be constructed regardless; if $a$ is larger, the administrative costs of compliance will discourage firms from undertaking the projects. In a very rough sense, the administrative expense forces firms to internalize some of the costs of their own pollution.

Now consider two firms within the same geographic vicinity, each of which is contemplating erecting a factory. The two factories have the same cost and produce the same pollution, but factory 1 provides greater benefits: $b_1 > b_2 > c$. Imagine that each factory by itself would be welfare enhancing: $b_1, b_2 > c + a + d$. If the costs of pollution are linear—in other words, if each factory produces pollution with cost $d$, and together they produce pollution with cost $2d$—then both projects are worth pursing. However, the costs of factory pollution may not scale linearly. Scientists suspect that many environmental responses to pollution behave non-linearly or accelerate when pollution levels cross a certain threshold. It is conceivable that the combined pollution cost from both factories would be not $2d$, but $2d + s$, where $s$ is some undesirable synergy created by the two pollution sources. Together, the factories may not be welfare enhancing.

78 For instance, the climatological response to greenhouse gas emissions is probably highly complex and non-linear in some domains. See Alley et al. (2003); Reilly et al. (2001). Dose-response relationships may also be non-linear in a variety of disparate contexts ranging from nuclear waste contamination (see Viscusi 2005, 235) to drug treatments. See Strnad (2005, 1229).
enhancing, if \( b_1 + b_2 < 2c + 2a + 2d + s \). In this case, welfare would be maximized if firm 1 constructed factory 1 (at a benefit of \( b_1 \)) and firm 2 decided to forgo factory 2.\(^79\) If \( a \) is sufficiently high such that \( b_2 \approx c + a \),\(^80\) this is exactly what will occur. The high costs of procedural compliance will act as a costly screen against the lowest-value polluters. In this type of situation, the costly screen exploits the asymmetry between the initial cost of the first factory and the additional cost of the second factory. By eliminating the lowest-value factory and avoiding the multiplying effects generated by a second polluter, the costly screen would generate benefits that exceed its costs.

The problem with high process costs as a passive barrier is that they are themselves likely to expend much of the surplus they create. Better, as in the patent context, to minimize administrative costs and replace them with high administrative fees, which are not deadweight losses and could be plugged directly back into national or state fiscs. Section 110 of the Clean Air Act already authorizes states to impose fees in association with permit applications,\(^81\) though few states have availed themselves of the opportunity and none imposes fees of the necessary magnitude. Alternatively, emissions trading regimes could be used to select for the highest-value polluters while simultaneously holding total pollution beneath any desired threshold (see Freeman & Farber 2005, 814–822). Such trading schemes remain rare, with the Clean Air Act’s sulfur dioxide deposition program standing as the only prominent national example.\(^82\) Under the right conditions, fees or emissions-trading programs might be profitably deployed. In their absence, process costs may offer a second-best solution.

\(^79\) This is a specific illustration of the more general point that negligence and regulatory rules cannot effectively control activity levels. See Shavell (1980). Where regulatory rules themselves will fail, the administrative costs of compliance with those regulatory rules may turn out to have some beneficial effects.

\(^80\) Of course, \( b_2 > c + a \) by assumption. However, uncertainty and risk aversion might dissuade firms from investing in a project that is not obviously cost-benefit justified.

\(^81\) 42 U.S.C. § 7475(a)(2)(A) (2007) (listing “economic incentives such as fees, marketable permits, and the auction of emissions rights” as tools available to state regulatory agencies).

4.3. Immigration and Naturalization

In a recent article, Adam Cox and Eric Posner suggest that the United States’ peculiar combination of methods of border control and naturalization function in large degree as an inducement to self-screening (Cox & Posner 2007). For example, the physical barriers to entry into the country, much like the administrative processes of having a patent examined and granted, are costly to overcome but nearly always surmountable (sometimes literally so). These “process” costs thus exist in part to force potential immigrants to reveal private information about their expected productivity within the United States (Cox & Posner 2007, 824–827). Only immigrants who believe that they will be able to earn a great deal of money—and thus the immigrants that, by one metric, are most desirable—will elect to attempt entry.

Related types of costly screening exist throughout the immigration system. For instance, consider an employer who wishes to hire a highly skilled foreign worker using an H-1B visa. H-1B visas are accompanied by a host of procedural requirements. These include, first and foremost, the filing of an extended series of documents with the U.S. Citizenship and Immigration Services (“CIS”) and the payment of fees totaling $2,190. In addition, the employer must certify to the Department of Labor that the H-1B visa holder is earning a salary commensurate with American workers performing the same jobs, and must “[p]rovide working conditions for [H-1b holders] that will not adversely affect the working conditions of workers similarly employed.”

Certain U.S. employers may, in addition, be classified as “H-1B dependent” if a particularly large fraction of their employees are H-1B visa holders. Every H-1B dependent employer must certify: (1) that its

---

83 The H-1B is a special class of visa available to non-citizens who work in “specialty occupation[s],” defined as an occupation that involves the application of a “body of highly specialized knowledge” and requires at least a bachelor’s degree. 8 U.S.C. § 1184(i)(1) (2007). The majority of H-1B recipients are scientists, engineers, doctors, and other technically trained professionals. See Ifill (2007, 504 n. 87).

84 Instructions for Completing Form I-129, available at http://www.uscis.gov/files/form/i-129instr.pdf, at 17–18. The fee is reduced to $1,440 if the employer is a small business. Id.


86 20 CFR § 655.736(a)(1) (2007). This fraction varies based upon the size of the company, and for companies with 51 or more employees it is equal to 15 percent. Id. § 655.736(a)(1)(iii)(A) & (B).
H-1B-holding workers will not “displace” any American workers; and (2) that it has attempted to recruit American workers to fill the open positions before hiring workers via the H-1B process. In practice, this means that the employer must advertise the open position in a newspaper or trade publication before hiring a foreign worker using an H-1B. This process is costly and can involve a significant amount of delay. Immigration-centered law firms typically charge in the neighborhood of $2,000 to complete an H-1B visa application, exclusive of the fees paid to the CIS and the cost of running a newspaper advertisement and determining the appropriate market wage. In total, then, an employer will typically spend in the neighborhood of $5,000 to hire a foreign worker using an H-1B visa.

The procedural requirements involved in the H-1B process—certification that no American workers are available, that the employer is paying the prevailing wage, and so forth—are ostensibly designed to ensure that the employment of H-1B workers will not redound to the detriment of American workers. These processes are a type of substantive examination: they force employers to collect and disclose the particular information that interests the government (and the public at large). At the same time, the screen is not a particularly effective one. Employers view the duty to advertise an open position as little more than a pro forma requirement, and tales of fraud and misconduct abound. The substantive examination may be worth little.

Simultaneously, though, the private cost of hiring an H-1B worker functions as a costly screening device that forces the employer to ascertain

88 Id. § 655.739 (2007); see also Form ETA 9035, available at http://www.dol.gov/libraryforms/forms/ETA/Form_ETA_9035.pdf, at 2.
89 See 20 CFR § 655.739(d)(2)(ii) (“Passive solicitation methods include advertising in general distribution publications, trade or professional journals, or special interest publications (e.g., student-oriented; targeted to underrepresented groups, including minorities, persons with disabilities, and residents of rural areas); America’s Job Bank or other Internet sites advertising job vacancies; notices at the employer’s worksite(s) and/or on the employer’s Internet ‘home page.’”).
90 This information was provided by Davidson & Schiller, an immigration law firm located in Chicago, Illinois. Notes on file with author.
91 See, e.g., The Scams & Problems of H-1B Visas, available at http://www.edu-cyberpg.com/Teachers/H1B.html (“The scam here is that they put little tiny ads in the San Jose Mercury News with almost all the words abbreviated and in the smallest type they can find.”).
and evaluate the same information that the government has targeted actively. Hiring a foreign worker in lieu of an American citizen arguably imposes some social cost upon the United States as a whole. At the same time, in many cases it confers a benefit upon the employing firm (and, by extension, upon the country at large) by supplying skilled labor for which there may be no American substitute. The harder it is to find a American worker to fill a high-technology job, the greater the private value and the social value of hiring a foreign H-1B worker. By forcing the employer to navigate a series of expensive administrative procedures before obtaining an H-1B visa, the government selects against those employers who would seek to realize only very small private benefits from hiring H-1B workers (when American workers are available) and in so doing might impose domestic social harms by refusing to hire qualified citizens. As in the patent context, the lineup of costs and benefits is likely asymmetric: situations in which it would be highly socially beneficial but only loosely privately beneficial to hire a foreign worker likely do not exist.

At the same time, and in similar fashion to the administrative contexts described above, many of the benefits will be consumed by the same transaction costs that catalyze the necessary screening. A simple substitution of higher fees for process would likely be welfare-enhancing. Nonetheless, the screening function performed by costly CIS and DOL procedures may usefully complement the operation of those agencies’ substantive examinations.

5. CONCLUSION

Patents do not come cheaply to applicants. Between the cost of hiring an attorney and the fees that an applicant must pay to the PTO, the average applicant spends approximately $22,000 to obtain a patent, and possibly much more if the patent is of debatable validity, concerns a complicated technology, or resides in a crowded technological field. Scholars have traditionally treated these expenses as nothing more than the purchase price of the patent examination, a series of costs to be avoided or minimized wherever possible. This approach has overlooked the fact that procedural costs function as a costly screen against low private/low social value patents, selecting disproportionately against this insidious class of property rights. The costly screen thus serves as a useful complement to substantive
patent examination, which has proven largely ineffective at preventing the accumulation of large numbers of worthless (and socially harmful) patents.

Administrative costs operate similarly in other contexts. The procedural costs involved in evictions and due process determinations may deter filings in which the action is worth little to the landlord or employer yet produces substantial social costs. The administrative expense involved in obtaining a NEPA permit may block factories that are not worth the pollution they will generate. And the administrative costs of hiring H-1B workers may select against those employers who would realize only very small private benefits from hiring H-1B workers while causing more substantial domestic social harms. Where substantive systems of examination are not fully effective, costly screens may play an important subsidiary role.

REFERENCES


