Private Profits and Public Benefits - How not to Reform the Patent System

Johannes Koenen*

University of Mannheim

May 9, 2010

preliminary version - all comments are warmly welcomed

Abstract

The need for a reform of the US patent system has been widely expressed, both by scientists and practitioners. There are widespread complaints about the quality of granted patents, especially the lack of "clearly drawn property lines", to the extent that some refer to the patent system as "failed". I present a model that includes the three pillars of current patent systems: legislation, patent and trademark offices and courts. I show that many of the contemplated (or demanded) reform steps, such as increasing the level of scrutiny exerted by the PTOs or increasing fees for patenting, may actually decrease the average quality of patent applications in relevant cases. Further, we find that the competitive setting plays a central role in determining whether courts can correct errors made by the PTOs, which raises empirically relevant questions.

*Financial support from the DFG is gratefully acknowledged. This article has greatly benefitted from helpful discussions with participants of the Innovation-Workshop of the SFB TR15 at the Reisensburg in Spring 2009. To contact the author: johannes.koenen[at]gmail.com.

1 Introduction

1.1 Motivation

A patent is not a property right. Patentees cannot directly exclude others from markets that are covered by the claims of their patent(s). Instead, a patent only conveys the right to go to court against alleged infringers to the holder. In recent years, the number of court cases involving patents has grown extremely rapidly to impressive levels.[1] In a thought-provoking study, Bessen and Meurer (2008b) estimate that litigation costs related to patents have exceeded private profits from patents ever since the late 1990s. According to their estimates, global profits directly connected to US patents in 1999 accrued to 9.3 billion USD, while their estimate for the domestic litigation costs to companies is roughly 16 billion USD. These figures are alarming.

The need for a reform of the US patent system has been widely expressed, both by scientists and practitioners. There are widespread complaints about the quality of granted patents, especially the lack of "clearly drawn property lines", to the extent that some refer to the patent system as "failed". Currently, there are efforts underway in both Europe and the US to improve the way patents are granted and enforced.[2] Researchers' more audacious recommendations for this have ranged from advising fundamental, game-changing reform such as Bessen and Meurer (2008a), to demanding the patent system as a whole being scrapped such as Boldrin and Levine (2002, 2008).

Compared to these demands, the actual reforms that have been attempted are extremely tame, focussing on (seemingly) minor procedural measures. And also the bulk of theoretical research has focused on more marginal shifts. The reason for this is fairly obvious - despite the many shortcomings of the current systems, the existing patents held by individuals and corporations are tremendously valuable. The investments that lead to the protected innovations were (arguably) sunk with the goal of patenting in mind, which means that any legislature tampering with the system will face never be-

[1] See Cook (2007) for an analysis of this phenomenon, with a focus of the role of specialized courts.

[2] Recent efforts in Europe include the development towards a unified "European Patent" agreed upon in the Council of Ministers of the European Union on December 4, 2009, as well as the introduction of a specialized European Patent Court as a part of the (currently stalled) European Patent Litigation Agreement. In the US, introducing new patent legislation has become a bi-annual tradition, of which the Patent Reform Acts of 2009, 2007, 2005 and 2003 bear witness. The main goal of the latest was to reduce the burden of the Patent and Trademark Office in the inspection process.

fore seen claims for damages. As the youngest current patents have another 20 years on the clock, any fundamental reform steps are all but certain to be postponed until then, at least.[3]

So for the foreseeable future, any revolution to the patent system by needs will be extremely marginal, infused with apparently harmless tweaks to the patent application fee or the responsible jurisdiction for patent claims; but in fact, due to the immense complexity of the current patent-machinery, pitfalls loom even for tweaks that appear to be harmless. The goal of this article is to provide a theoretical framework that is as simple as possible, while it still spans all three pillars of current patent systems. Our aim is to demonstrate a number of perhaps unexpected tradeoffs involved in tampering with them. What we consider the three pillars are:

(1) Patent legislation, which defines the prerequisites for a patent to be granted as well as the right of patent holders.

(2) Patent and trademark offices (PTOs), which apply the existing rules and regulations in examining patent applications.

(3) Courts, which uphold the rights of patentees against alleged infringers and, by giving third parties the chance to challenge existing patents, have the ability to invalidate patents that should not have been granted.

The basic dilemma in patenting is that the patentability of an idea from the perspective of legal requirements or social welfare does not necessarily coincide with the private benefits an inventor obtains from being granted patent claims. In our model, we distinguish between the profit an individual or firm may derive from an idea if patented and the patentability of an idea from the legal perspective. Unlike the existing literature, which mostly focuses on binary *good vs. bad* distinctions, we allow for a continuum along both dimensions - ideas may range from patentable over "almost" patentable to "clearly not" patentable, as is arguably the case in reality. It is the role of the PTO to make this distinction, with its decision necessarily being imperfect - as PTOs face

[3]With this point in time moving along until any first steps are at serious reform are taken. And the 20-year timeframe is actually rather optimistic, because research programs that are currently underway with patents potentially looming years from now may already have generated claims based on the investors' trust in the current system.

time, budget and staffing constraints. We represent this in a highly reduced form in the model. Finally, there is the possibility of patent holders and competitors becoming entangled in court proceedings, where we can distinguish between damage suits against alleged infringers on the one hand and challenges of the validity of granted patents on the other. We attempt to represent both kinds of proceedings in our model.

The occurrence, type and outcome of court interactions depend on various factors: The most important are the average quality of patents, as there is asymmetric information about the validity of *individual* patents, and the competitive setting between patentees and potential infringers. We show that one of the most frequently contemplated (or demanded) reform steps, increasing fees for patent applications at the PTO, may actually *decrease* the average quality of patent applications. Further, we point out some unintended and potentially costly consequences of increasing the level of scrutiny exerted by the PTOs. Further, we find that a reduction of litigation costs for patent costs has multiple desirable effects - which is especially relevant in the face of specialized courts of appeals being introduced in the US and the European Patent Legislation Agreement.

The remainder of the article proceeds as follows: After a brief review of the related literature, section 2 sketches legal and empirical differences between the European and the US patent system in order to derive stylized facts and motivate the relevant determinants of our model. Section 3 develops the model of patent legislation, examination and litigation. In section 4 we derive the equilibria of the model and perform comparative statics exercises to determine the effects of various approaches to patent system reform. Section 5 emphasizes empirically relevant predictions, proposes venues for further research and concludes.

1.2 Related Literature and Contribution

The first strain of the literature that our article is related to studies the effects of patent litigation. The issue of potentially invalid patents and resulting litigation has garnered recurring interest in the past two decades. Meurer (1989) proposes the first model that explicitly takes the possibility of patent invalidity and resulting litigation into account. Patentees face a single competitor, with both parties aware of the fact that courts may overturn the patent. The competitor can either accept the patentees non-cooperative bargaining offer, do nothing or challenge the patent through the courts. The study carefully shows how settlement and litigation probabilities depend, among other things, on the strength of the patent both under symmetric and asymmetric information. We adapt the model for one of our litigation subgames.

Crampes and Langinier (2002) study the opposite competitive setting: Here, the competitor does not challenge the patent, but simply enters the market, thereby (potentially) infringing the patentee's rights. The patentee does not necessarily notice this, instead he must invest in a costly monitoring technology, which determines the probability with which infringement is observed. If observed, the patentee must make the decision whether to accommodate, litigate or settle with the infringer, where the settlement results from cooperative Nash Bargaining. The otherwise efficient bargaining potentially breaks down as the authors impose costly frictions in the bargaining process. Monitoring is shown to be an effective way to preclude entry. Bessen and Meurer (2006)'s approach is very is similar to Crampes and Langinier (2002) - they give the potential infringer the strategic opportunity to invest into research, which affects the probability of being fined by courts in two possible ways: This activity could be general R&D (e.g. adding additional features to the product), which would potentially increase the risk of being found infringing. Or the company could exert effort to specifically invent around existing patents, which would reduce the probability of infringing. Instead of monitoring activity as in Crampes and Langinier (2002), the patent holder can invest in strengthening his patent; the probability that the court will find the infringer guilty if a trial arises increases in this investment (this can be interpreted as applying for additional patents to create a thicket or employing better and more expensive lawyers to phrase the patent application and claims optimally). In the empirical part of their project, the authors find that competitors' investment is overwhelmingly focused on research.

More specific issues within the area of patent litigation have also been discussed: A number of papers investigates the role of different liability rules affects the level of protection granted by intellectual property rights, see e.g. Schankerman and Scotchmer (2001) and Choi (2006) for comparisons of lost profit and unjust enrichment rules in different competitive settings. Anton and Yao (2007) study the infringement decision in the case of process innovations for lost profits damages. Antitrust issues and anticompetitive effects of the settlement of patent litigation are taken into account in Shapiro (2003) and Lemley and Shapiro (2005).

A second group of articles is concerned with the organization of and incentives within patent offices and their effects on the patent system. Caillaud and Duchene (2006) scrutinize the capacity of the patent office to deal with applications and the so called overload problem. In their model, if the PTO were to employ a strict examination standard, a hypothetical equilibrium exists in which bad applications are deterred. A problem arises, though, if the PTO is overloaded with applications - as it can no longer enforce the strict standard, deterrence becomes less effective and the separating equilibrium can no longer be upheld. In Prady (2008), low quality inventors can induce shirking of the patent examiner by sending signals that require more effort to disentangle. Schuett (2009) models the examination process as a combined moral-hazard/adverse-selection problem and tries to explain and argue for different incentive schemes for examiners.

Two recent studies are closely related to ours. Both Farrell and Shapiro (2008) and Chiou (2008) present models which allow for patent examination and potential litigation. In Farrell and Shapiro (2008), an upstream innovator owns a patent that may be "weak" in the sense that a court will only uphold it if challenged with a certain probability. The upstream innovator faces a set of downstream firms which can apply his innovation. Depending on the competitive setting, they show that even very weak patents can have a strong price-shifting effect – when the downstream market is very competitive, then there is next to no private incentive to challenge the patent, which gives even weak patents great power. In such a setting, more stringent PTO reviews are welfare enhancing. Our study is in many ways complementary to this approach: While in Farrell and Shapiro (2008) there is no room for litigation in equilibrium, despite zero litigation costs, our main focus is to determine under which conditions which kind of litigation will appear, and how this is affected by contemplated reforms. Their main analysis focuses on the role of the intensity of downstream competition in a specific setup - we treat this factor in a reduced form, which to a certain degree encompasses their analysis.

Motivated by Lemley (2001)'s provocative thesis that increases in the diligence of the patent office are inefficient and one should let the competition sort out bad patents through litigation, Chiou (2008) sets up a two stage model, in which first the patent office exerts effort in order to find prior art that allows to deny a latently invalid patent. If the patent office is unable to destroy the patent, in the second stage a private competitor can exert effort in the same way to have it repealed. While for relatively good patents (high priors of patent quality) he finds a crowding out of private efforts through the patent office along the lines of Lemley (2001), for low-quality patent population he finds the opposite effect: Stricter enforcement by the patent office actually encourages the competitor to himself invest in research efforts to try to invalidate the patent. The paper relies on a new way of modeling court interactions – higher exertion of (costly) effort by the challenger (or by the PTO) leads to the destruction of the patent with a higher probability and the "quality" of a patent affects the marginal effect of effort. Again, our approach, using a more classical way to model suits and countersuits applied to a population of patent applications, can be seen as complementary.

This article contributes to both strands of the literature. Our model encompasses both the examination and the litigation phase of the patenting process. Regarding litigation, we integrate the competitors' choice between challenging an existing patent or entering the market directly and potentially infringing. Competitors' beliefs concerning patent validity are determined by the policy adopted by the patent office as well as settlement offers proposed by the patent holders. Further, we allow for a delay between patent application and the final decision of the patent office - a form of patent office overload related to the one studied by Caillaud and Duchene (2006), which can be considered a "cost" of demanding higher levels of diligence from patent examiners and which awaits empirical exploration. Finally, we study a population of ideas which varies continuously regarding private value to the inventor and objective patentability, which allows us to shed light on tradeoffs due to the composition of the population of patent applications, which have not been studied before.

2 A Brief Comparison of the US and the European Patent System - Stylized Facts

Before we describe the theoretical model, we first provide a glimpse of some of the differences between two of the most important patent regimes today: the US and the European patent system. The goal is not to exhaustively display and analyze these issues, but to motivate the factors and variables we integrate into our model.

2.1 Patent Legislation and Examination through Patent Offices

Even though there are many calls for and initiatives in the direction of harmonization, the US and the European patent legislations retain some very distinct features. Generally speaking, there are stark differences between European and US patent laws, while the differences among the core countries of the EU are comparably negligible.

While the “first to file” system is prevalent in Europe, that is, the first to apply for a patent is awarded the right, traditionally the US award patent rights to the first person to discover an innovation (“first to invent”). A further difference regarding the patenting requirements concerns what in the US is termed “novelty” and, e.g., in Germany the “inventive step”. Both requirements state that for an idea to be patentable it may not yet have been made public previously in any form. In Germany this holds absolutely. In the US, inventors are granted a ”grace period”, i.e. they generally only have to file their application within a year of their idea’s publication.

The other general requirements for something to be patentable are similar in both systems - patentable subject matter, non-obviousness, and applicability. Still, there is quite a lot of evidence that these requirements are interpreted differently by the US and European patent offices, which is close to the focus of our study. For example, Straus and Klunker (2007) cite the following numbers of the US and the European patent offices: In 2005, there were 409,532 applications for patents in the US. Out of these, 165,485 were granted, i.e. a share of 40.4 per cent. On the other hand, in Europe, the number of applications in the same year was 197,391 with 53,256 patents being granted, i.e. a significantly smaller share of 27,0 per cent. From these numbers alone, it appears that that the German patent offices apply a stricter standard than the American ones. This is corroborated by other observations: The average time from filing a patent to it being granted was 45,3 months in Europe, while it was only 24 months in the US, as Hall and Harhoff (2004) report. They also look at the grant-rates of patents at the European Patent office for US patents seeking European approval on the one and patents from other countries seeking approval in Europe on the other hand. They find that the approval rate for US patents was substantially lower, with a 16 per cent difference in 1995. The trend points towards a deterioration of the US standards, as in 1979 there was parity concerning these numbers.

Straus and Klunker (2007) further cite a study by the consulting firm Roland Berger that found that in order to apply for and maintain a patent over its entire lifetime, costs between 32,000 and 47,000 EUR accrue in Europe, while the average figure is only 10,250 EUR in the US.[4]

From these observations, we derive the following stylized facts regarding the patent application process in the US and in Germany:

(F1) Patent applications are scrutinized more strictly by European than by US patent offices.

(F1') As a result, a European patent is potentially a better indicator of the strength of the patentee's claims than a US patent.

(F2) It is significantly more expensive to obtain a patent in Europe than in the US in terms of the fees required.

(F3) The duration of patent pending varies between patenting systems by a considerable margin.

2.2 Courts and Patent Trials

After briefly discussing the differences concerning the laws in the two jurisdictions, we will next consider how patent suits fare in the courts.

Cook (2007) notes that in the course of the last decade, the number of patent cases filed in the US has roughly doubled, from about 1,250 in 1990 to about 2,500 in 2000. He shows that this is one the one hand a result of more patents being filed in general (the ratio of cases per granted patent remains relatively unchanged, as the latter number increases from about 90,000 to about 180,000 per year in the same period). On the other hand it results from courts being more accommodating to plaintiffs. In his empirical model the decision to go to court depends, among other factors, on the share of cases that were successful in the respective district in previous periods. One impressive measure in this regard is that the probability of receiving a patent reward of 1 million USD or more in 2001 constant dollars increased from less than 10 percent in 1976 to more than 30 percent in 2000.

[4]For comparison: Regular period from filing to grant: 31,6 months in Japan, 45,3 months in Europe, 24 months in China and US. Patent costs are estimated to range from about 2,400 to 4,000 EUR in China.

Concerning the average costs of going to court over patent infringement in the US, estimates range from 500,000 USD to 3,000,000 USD for each party. Much larger sums are mentioned in the context of complex or high-stakes cases, especially in the area of pharmaceuticals. For a European patent to be litigated, court fees of 70,000 EUR arise according to European Patent Office (EPO) reports. Their estimates of the lawyers' fees bourn by the parties in addition amounts to approximately the same figure, so that each party would have to expect costs around 150.000 EUR ex ante.

We glean the following stylized fact from these brief observations:

(F4) There is a significant cost difference between bringing a case to court in the US and Europe, i.e. the costs of litigation are vastly higher in the US than in Europe.

3 Model Setting

3.1 The rules of the patent system

Legislators (exogenously) define a minimum standard for utility, novelty, non-obviousness, etc which makes an idea objectively patentable. For this model, we assume that these various criteria can be reduced to one dimension, and the minimum objective standard set by legislators is denoted μ.

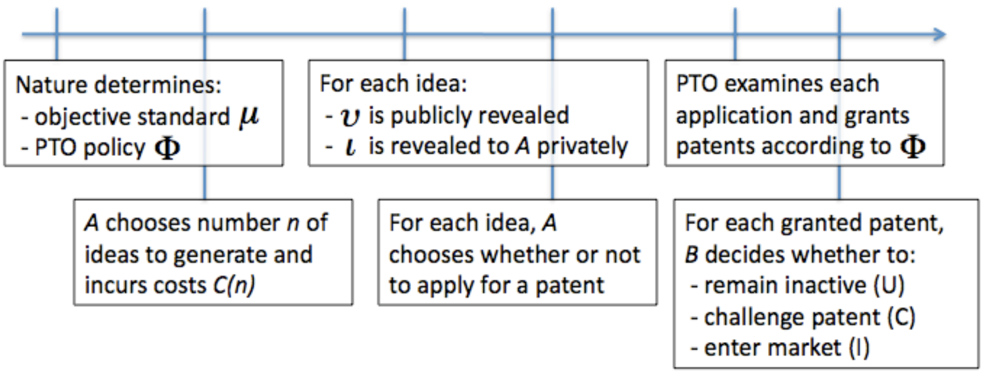

Figure 1: Timeline of the entire model

3.2 Strategic Players and "Ideas"

We consider a game with two risk-neutral strategic players, A and B. Player A actively generates ideas. Ideas are "random events", i.e. individual ideas are imbued with certain randomly drawn characteristics (more on this below). Generating n ideas in a given period, A incurs costs $C(n)$. We assume that $\frac{\partial C(n)}{\partial n} > 0$ and $\frac{\partial^2 C(n)}{\partial n^2} > 0$.[5]

Each individual idea that is generated is defined by the following characteristics:

(1) The patentability of the idea from a legal standpoint, i.e. its utility, novelty, non-obviousness etc. Again we assume that these can be reduced to a single scalar, which we denote by ι, with $\iota \epsilon [0, 1]$. Note that if $\iota \geq \mu$ an idea is objectively patentable from the legal standpoint.

(2) A measure of the additional value to the inventor derived from patenting the idea, which we denote by υ. To be more specific, this measure is defined as the value that the inventor can appropriate from exploiting the idea optimally as a monopolist after obtaining the patent, i.e. given no-one infringes upon the patent. As we are concerned with the incentive effects of patents in the face of challenges with respect to enforcement, we set the value from exploiting an idea without patent protection to zero. Therefore in our model, from an ex ante perspective, an idea is only valuable to its inventor in as far it can be patented. We normalize once more such that $\upsilon \in [0, 1]$.

The inventor only learns the actual characteristics of a given idea *after* it has been generated and he has incurred the associated costs. Specifically, the parameters ι and υ are respectively drawn from commonly known and independent[6] distributions $F(\iota)$ and $G(\upsilon)$ and revealed to A. We further assume that F and G are both continuous and strictly increasing.

[5] This can be interpreted in the following way: A resembles the population of potential inventors and they are sorted according to the marginal costs of their ideas.

[6] We only use this assumption (independence of distributions) to be able to derive a closed form solution for an expression below. The majority of our results hold without it.

3.3 Patent and Trademark Office

3.3.1 Patent Applications and the Patent Office

A decides for which of his ideas to submit patent applications. It costs A the fixed amount τ_1 to submit an idea to be examined by the patent office. We assume that the PTO can only make a binary decision regarding a given patent application: either approve or reject it. According to the legal rules stated above, the PTO should approve patent applications if and only if $\iota \geq \mu$. We assume that the PTO cannot directly observe the actual ι of an individual application. Conceptually therefore, it could commit two kinds of errors: Errors of type one, i.e. granting a patent despite $\iota < \mu$, and errors of type two, i.e. declining patents despite $\iota \geq \mu$ and both of these errors are associated with social costs. For the sake of this model, we neglect errors of type two - discussions both with scientists and practitioners have convinced us that they are empirically close to irrelevant, since there is a relatively fast and cheap way to appeal the decision in both jurisdictions in this case. Imagine that A can incur the additional (small) costs τ_2 in order to have the application reexamined which leads to it being granted if $\iota \geq \mu$. The only effect of this will be to shift the expected profits from patenting of holders of patentable ideas downwards by a fixed sum.

We model the patent office policy in the most simple possible reduced form. The patent office simply implements an exogenously given examination policy Φ which determines the likelihood $\phi(\iota)$ of a given idea being patented, including the probability of a first order mistake being made if $\iota < \mu$. For example, if the PTO grants every patent application, then $\phi(\iota) = 1$ for any ι. The inspection policy does not depend on υ – for one, patent examiners (as everyone else) have a very hard time determining the expected value of a patent, and also, more importantly, guidelines in general forbid different treatment of applications depending on their suspected value. We assume that $\phi(\iota) > 0$ for all ι, i.e. even the worst quality idea always has a positive chance to be awarded a patent. Further we assume that $\frac{\partial\phi}{\partial\iota} > 0$ if $\iota < \mu$, i.e. the closer ι comes to the objective patenting threshold, the more likely a patent is to be (falsely) granted. Therefore the probability with which a given application will be granted is $\phi(\iota) \leq 1$ if $\iota < \mu$, and $\phi(\iota) = 1$ if the idea is objectively patentable.

[Consider the following example for a possible micro-foundation as an illustration of the idea: The patent office generates a number ζ of signals, where ζ is part of the

exogenous schedule Φ. ζ *can for example be interpreted as the time that a patent examiner spends researching prior art or perusing the patent application. If individual signals are normally distributed with mean* ι *and standard deviation* σ, *the mean of the signals* $\bar{\iota}$ *has the expected value* ι *and the standard deviation* $\varsigma = \frac{\sigma}{\zeta^{0.5}}$. *The examination schedule further includes a cutoff level* μ^*. *If the mean of the generated signals is below this cutoff level, the PTO rejects the application, and it grants the patent otherwise.* ζ *and* μ^* *therefore implicitly determine the probability of type one and type two errors that the PTO commits for any given* ι. *Denote the normal distribution with mean* ι *and variance* ς *as* $N_{\iota,\varsigma}$. *Then the probability for a type 1-error given that* $\iota \geq \mu$ *is simply* $N_{\iota,\varsigma}(\mu^*)$, *and equivalently the probability for a type 2-error (which we do not consider) given* $\iota < \mu$ *would be* $1 - N_{\iota,\varsigma}(\mu^*)$.*]*

3.3.2 Extension: Period of Patent Pending

As discussed above, the decision whether or not a patent application will be granted is by no means made immediately. Instead, a substantial period of time passes between application and decision, during which the ideas has the status of "patent pending". Little investigation has focused on the value of patents in the course of the application process so far. But when Steve Jobs exclaimed his now famous “And, boy, have we patented it!” during the speech introducing Apple’s iPhone at the MacWorld Expo in January 2007, the greatest share of the patents protecting the touch-screen technology involved *had not been granted, yet.* In fact each of the 21 iPhone-patents considered “central” by technology-afficionadoes was still pending at this point in time.[7] Part of this stance was surely justified by the knowledge that the patents would be granted, later on, which turned out to be true. But to some extent, pending patent applications itself are valuable, especially in fast-moving industries.

There are a number of reasons to suspect that inventors do benefit from their ideas during the pending period. Many if not most license agreements are negotiated prior to the actual patent being issued. Other firms may think twice about entering the market if the product is designated with the "patent pending" stamp, fearing future lawsuits. Finally, in one of the few empirical studies related to the topic, Häussler, Harhoff, and Müller (2009) find that the chance to receive venture capital rises with the number of

[7]See http://www.mad4mobilephones.com/the-21-most-important-iphone-patents/562/ for an extremely detailed description and links to the individual claims.

patent applications in the firms' portfolios.

Again, we reduce this concept as far as possible. We simply denote the patent-pending value of an idea (ι, υ) as $\delta(\Phi)\upsilon$, where $0 \leq \delta \ < 1$ and $\frac{\partial \delta}{\partial \phi} < 0$. Therefore increasing the quality of patent office scrutiny (or lowering $\phi(\iota)$ is "expensive" in the sense that it increases the patent pending value of ideas, for example by increasing the duration between application and examination on average. We will not over-stretch this concept in the following, especially in this overly simple linear specification, but it is useful to remember this factor when discussing reforms aimed at more stringent inspection policies through the PTO.

[Consider again the illustrative example above. Imagine that the patent pending duration of a given idea after application is t periods, where $t = t(\zeta, n)$ and $\frac{\partial t}{\partial \zeta} > 0$. During this time, the perspective patentee receives the benefits $\Delta(\upsilon, \iota)$ per period.]

3.4 The Court System

We have described above how ideas are generated, how A decides whether or not to apply for patents, and whether or not patents are granted by the PTO. In the next and final stage of the model, B enters the picture. We consider B to be the only strategically acting competitor of A in each of the markets covered by a patent in the economy.[8] B observes which patents were granted to A and chooses one of the following reactions: He may either leave the corresponding market uncontested (U), challenge the validity of the patent before court (C) or enter the market and thereby infringe A's patent (I).

Choices (C) and (I) respectively induce different litigation games: For (C), we adapt the approach introduced by Meurer (1989). For choice (I), we use a setup related to Crampes and Langinier (2002), but with non-cooperative bargaining.

3.4.1 Information

While the patent holder is perfectly informed about the quality (ι, υ) of any given patent, B cannot observe ι. This parameter has the following significance in the context of court

[8] Analogously to above, B can be interpreted as the population of potential competitors to the population of inventors A. More specifically, B is the most efficient or profitable competitor of A - this mitigates the loss of generality from assuming only one competitor, as other less efficient firms' incentives are aligned with B's, yet less strong.

proceedings: The probability that a court upholds a given patent obviously depends on the legal patentability of the underlying idea. For simplicity, we assume that courts make the "right" decision, that is, they uphold the validity of the patent whenever $\iota \geq \mu$. One might question the assumption that courts can discover the true ι of an idea, while the PTO cannot. But seeing that court cases take years while patent examiners have days (if that!) to come to a decision regarding a given application, the assumption may appear less severe.

Therefore, from A's perspective, there is no uncertainty regarding the court's decision on the validity of a given patent: if $\iota \geq \mu$ the court will uphold its claims, and it will repeal the patent otherwise. As B cannot observe ι, though, the court proceedings take on a random character from his point of view and he must form the subjective probability of a patent being upheld given the information available to him. In our simple model, B only observes three things: First, the fact that a patent was granted by the PTO. Second, the value υ of the patent to the patent holder, which he may e.g. derive from observed sales. Finally third, eventual settlement offers S proposed in the course of the litigation game by A. From these observations, B forms the conditional subjective expectation $p_B(\iota \geq \mu | \upsilon, \Phi, S)$ of the validity of the patent. We will abbreviate the probability unconditional on a settlement offer S as p_B in the following and denote updated equilibrium beliefs following an informative signal S as p_B^*.

3.4.2 Structure of the Post-Patenting Litigation-Game

Combining the ingredients above, as the final step of the model, for each patent that was granted to A, the following subgame is played:

B initially decides whether to leave the patent uncontested (U), challenge the validity before a court of law (C) or enter the market and infringe A's patent (I).

(U) If B leaves the market uncontested, the game regarding the given patent ends and the players get the payoffs $\pi_U^A = \upsilon$ and $\pi_U^B = 0$.

(C) This case is treated analogously to, for example, Meurer (1989) and Chiou (2008). If B announces her intention to challenge the patent, A and B first have the opportunity to negotiate a settlement. The informed party, i.e. A, who knows whether or not the patent will be held valid by the court, can first propose a settlement offer S_C. B observes the settlement offer and in this process updates her subjective probability that the patent is valid. Then she decides whether or not to go to court, with the following outcomes: If $\iota \geq \mu$ the court upholds the patent despite the challenge and the payoffs are $\upsilon - \kappa$ for A and $-\kappa$ for B. If $\iota < \mu$ the court rules that the patent is invalid and therefore the market is no longer protected. By definition, the payoff of A therefore is $-\kappa$ and B obtains the payoff from a competitive market with free entry net of court fees which we denote as $\upsilon_B^C - \kappa$.[9]

(I) The following subgame ensues if B unilaterally enters the market with a product that potentially infringes A's patent. We assume that (C) and (I) are mutually exclusive, i.e. it is impossible to, for example, challenge a patent *and* enter the market.[10] Again, we assume that it is the informed party A who makes the settlement offer S_I. If B rejects the offer, A decides whether or not to litigate. As the patent is either valid ($\iota \geq \mu$) or not valid and A is perfectly informed about this, the case in which he sues and the courts invalidate the patent does not arise (here, A would receive a payoff of $-\kappa$ and B receives the competitive payoff net of court fees $\upsilon_B^C - \kappa$; therefore it is better for A to accommodate given patent invalidity, which gives the duopoly payoff to both parties). If the patent is valid and A sues, the courts grant him lost profits or compensatory damages, therefore his payoffs from litigation are $\upsilon - \kappa$.[11] On the other hand, the payoff of the infringer in this case is her duopoly profit plus the difference between A's duopoly profit and the forgone monopoly profit minus court costs, i.e. $\upsilon_B^D - (\upsilon - \upsilon_A^D) - \kappa$.

[9] Note that any market payoff always depends on the overall profitability of the market, therefore for example υ_B^C is actually $\upsilon_B^C(\upsilon)$. We continue to use the former as a shorthand.

[10] There are a number of sound economic reasons underlying this assumption. For one, challenging a patent before introducing a product raises the specter of willful infringement and punitively higher compensation in the case of being found infringing - we leave this consideration out of the model. Further, if the company is not found to be infringing, it may itself benefit indirectly from the patent protection of the market, as we will see below.

[11] We only focus on this form of damages, for an excellent discussion of the effects of different kinds of damage awards see for example Schankerman and Scotchmer (2001) or Choi (2006). Further, we abstract from the fact that an alleged infringement of a valid patent is not necessarily covered by the claims of the patent. One could simply rescale the payoffs of the players by the ex ante probability that the product will be found infringing, without significant changes to the results. To reduce the notational burden, we abstain from this exercize.

3.4.3 The Structure of Industry Payoffs

We make the following assumptions with regard to the industry profits. First, we assume that $\upsilon > \upsilon_A^D + \upsilon_B^D$, the profit that A can optimally extract is larger than the sum of the duopoly profits of A and B.[12] Second, we assume that $\upsilon_B^D \geq \upsilon_B^C > 0$. B makes weakly higher profits in duopoly than under full entry competition, but still makes strictly positive profits in the latter case (though they may be arbitrarily close to 0). Third, we assume that υ_A^D, υ_B^D and υ_B^C are all continuous and increasing in υ - loosely speaking, a *more* valuable market in general leads to higher duopoly and competitive payoffs for market participants. Each of these would hold for generic specifications of Cournot competition with entry under integer constraints, for example. Fourth and finally, we assume that $(\upsilon - \upsilon_B^C)$ and $(\upsilon - \upsilon_A^D)$ are nondecreasing in υ. As the market as a whole becomes more valuable, the differences between monopoly- and duopoly-, or monopoly- and competitive profits do not grow smaller.

4 Analysis

4.1 Equilibria

In the following, we first derive the Bayes-Nash Equilibria for the litigation subgames, before embedding them in the larger model in order to learn more about the incentives to generate ideas and patent them.

4.1.1 Litigation Subgames

We first focus on the case that has already been studied in the literature, i.e. **subgame (C)** in which B threatens to challenge A's patent before court.[13] Clearly this threat is not credible unless the unconditional expected payoff from going to court is non-negative,

[12] Implicitly, we thereby disregard such phenomena as "patent trolls", companies which cannot exploit their patents effectively *without* going to court (or threatening to do so) against producers who inadvertently infringed upon their patent. See Reitzig, Henkel, and Heath (2007) for a study focusing on this topic.

[13] As the derived results for this subgame have been established previously, we cover this case very briefly. For a detailed discussion of the technical aspects of the equilibrium, we refer to Meurer (1989).

i.e. the following condition holds.

$$(1 - p_B)\upsilon_B^c - \kappa \geq 0 \quad (1)$$

If (1) is violated, neither holders of valid nor of invalid patents make a settlement offer to B, who in turn does not go to court over the issue. This resembles a pooling equilibrium with the payoffs $\pi_A = \upsilon$ and $\pi_B = 0$.

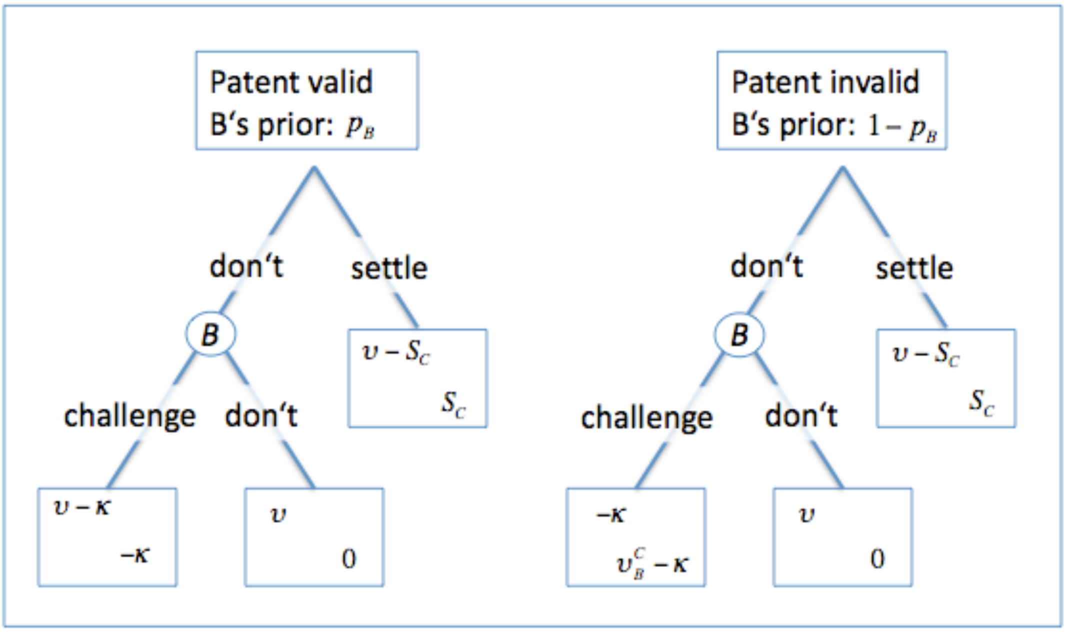

Figure 2: Game Tree for the Challenge Litigation Subgame

The clearly more interesting case is the one in which (1) is satisfied. Note that if $\iota \geq \mu$, A cannot credibly signal to B that he is the owner of a valid patent via his settlement offer and keep him from going to court completely, as signals are costless and can be mimicked by holders of invalid patents. Intuitively, no matter which sum holders of good patents offer to their competitor, they can always be mimicked by the holders of bad patents and litigation will occur with positive probability in equilibrium as a result. The following is an equilibrium: Holders of good patents make the toughest settlement offer possible, i.e. $\hat{S}_C = 0$. Holders of bad patents mix between $\hat{S}_C$ and offering a settlement worth $\underline{S}_C = \upsilon_B^C - \kappa$.

Lemma 1: *[Proposition 1 - Meurer (1989)] Perfect Bayesian Equilibrium of subgame (C): (i) If (1) is violated, in equilibrium no settlement offer is made and no patent is* *challenged. (ii) If (1) is satisfied, a semi-separating equilibrium arises in which holders of good patents never make a settlement offer, holders of bad patents mix between making no settlement offer and offering $S_C = \upsilon_B^C - \kappa$ and B updates his beliefs to p_B^* upon receiving or not receiving a settlement offer and mixes between litigation and inaction.*

We briefly sketch the existence proof in the following in order to clearly demonstrate the mechanics and intuition of the equilibrium. With the refinements of sequential equilibrium and D1, this equilibrium is also unique - for these proofs we refer the reader to Meurer (1989).

Proof. (i) As the expected payoff from litigation given p_B is negative and no signal is forthcoming, no updating takes place and B chooses inaction. (ii) Denote the probability with which B goes to court given the signal $S_C = 0$ as λ. Then the expected payoff of low quality patentees from mimicking the behavior of high quality patentees is: $(1-\lambda)\upsilon + \lambda(-\kappa)$, while the certain payoff given the settlement offer is $\upsilon - (\upsilon_B^C - \kappa)$. For low quality As to be willing to mix between these two, they have to be identical. Consider next the payoff of B if she observes the signal $S_C = 0$. Let us call the probability with which low-quality patentees mimic good quality patentees β. Upon perceiving the signal $S_C = 0$, Bayesian updating of the prior gives us the following condition: $1 - p_B^* = \frac{\beta(1-p_B)}{p_B + \beta(1-p_B)}$. For B to be willing to mix between litigation and accepting the tough settlement offer, the following equality has to hold: $p_B^*(-\kappa) + (1-p_B^*)(\upsilon_B^C - \kappa) = 0$. For the following values of λ and β, all of these conditions are fulfilled simultaneously: $\lambda^* = \frac{\upsilon_B^C - \kappa}{\upsilon + \kappa}$ and $\beta^* = \frac{p_B}{1-p_B}\frac{\kappa}{\upsilon_B^C - \kappa}$. Finally, specify off the equilibrium path beliefs $p_B(\iota \geq \mu | \upsilon, \Phi, S > 0) = 0$. Then it is easy to show that no profitable deviation exists for either player.

Note that litigation arises here as a result of the competitor not being able to distinguish between good and bad patents given that the offered settlement is 0, while she can identify those patent holders as bad who offer a more generous settlement. For these, though, B needs not litigate, as she receives the same expected payoff from the settlement as she would from litigation. Note that the holders of bad patents are exactly indifferent between the tough settlement offer of 0 and the generous offer due to the chance that the competitor will take the case to court with a positive probability given the tough offer, which results in them losing their patent protection entirely.

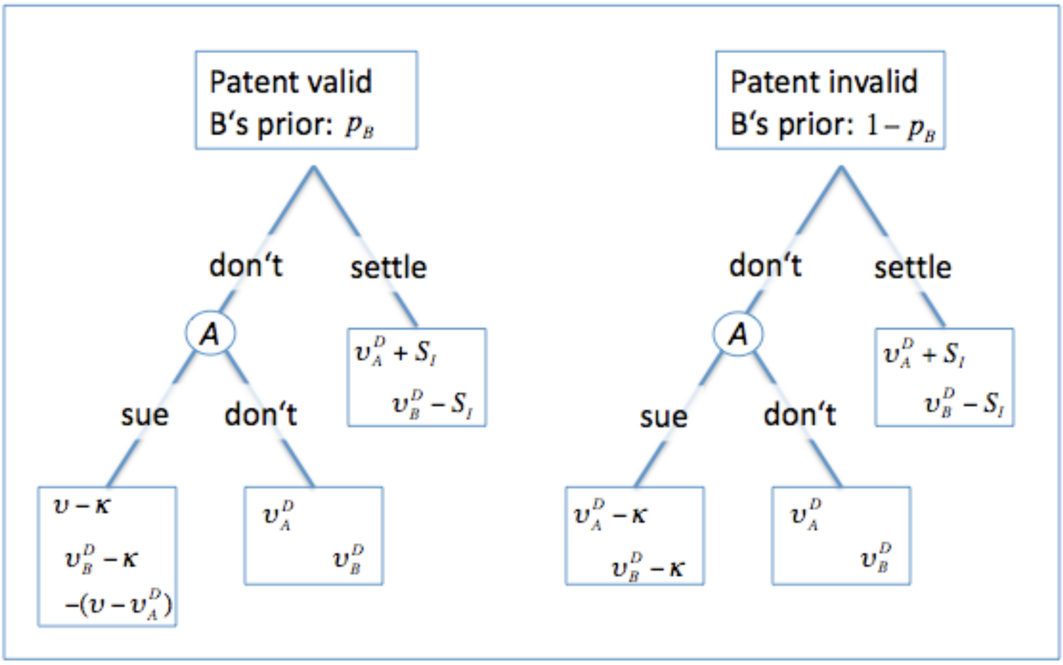

Figure 3: Game Tree for the Infringement Litigation Subgame

Let us next consider the **subgame (I)**, in which B enters the market, thereby infringing the potentially valid patent of A. From the payoff-structure above it is clear that it is never profitable for holders of bad patents to go to court, as the court is certain to deem their patent invalid and clearly $v_D^A > -\kappa$. Holders of valid patents will prefer litigation to inaction only if the (certain) court outcome of monopoly profits net of court costs is larger than the duopoly outcome they would receive given inaction, i.e. the following condition holds:

$$v - v_A^D \geq \kappa \tag{2}$$

If (2) is violated, even the patentee with a valid patent cannot credibly threaten to go to court, therefore B will not be willing to pay any compensation in settlement negotiations and $S_I = 0$. The payoffs of the two parties are $\pi_I^A = v_D^A$ and $\pi_I^B = v_D^B$. Again, the more interesting case is when (2) is satisfied. Here, it is profitable for holders of good patents to sue for infringement damages, while it is still not in the interest of holders of bad patents to do so. But as opposed to subgame (C), now it is the *informed* party who decides whether or not to go to court. Unlike in the previous case therefore there is no positive probability of litigation to induce patentees of the bad type not to mimic the good type and even a semi-separating equilibrium cannot be supported. Therefore in equilibrium, no informative settlement offer exists, in the sense that it allows B to update her prior probability of patent validity.

Good patentees know that they will receive a gain of $(\upsilon - \kappa) - \upsilon_A^D$ from litigation over inaction. The expected loss of B from failing to reach a settlement, on the other hand, is equal to the expected damages payment plus court costs, i.e. $p_B(\upsilon - \upsilon_A^D + \kappa)$. From this we derive the following condition:

$$p_B(\upsilon - \upsilon_A^D + \kappa) \geq (\upsilon - \kappa) - \upsilon_A^D \tag{3}$$

If this condition is violated, then holders of valid patents and the competitor are unable to obtain a settlement, and the patent holder sue for damages. These considerations allow us to formulate the following lemma:

Lemma 2: *Perfect Bayesian Equilibrium of subgame (I): (i) if (2) is violated, no settlement is reached and A remains inactive in equilibrium. (ii) If (2) and (3) are satisfied , settlement is reached with certainty. (iii) If (2) is satisfied and (3) is violated, no settlement is reached. Upon failure, holders of valid patents sue and holders of invalid patents remain inactive.*

Proof: (i) When (2) is violated, the threat of litigation is not credible for either kind of patentee, therefore no non-negative settlement demand will be met. (ii) Define a settlement offer $S_I^* \epsilon [(\upsilon - \kappa) - \upsilon_A^D, p_B(\upsilon - \upsilon_A^D + \kappa)]$. Then the following pair of strategies is a Perfect Bayesian equilibrium: Both types of A offer S_I^*. B accepts S_I^* and declines any other settlement offer. If a settlement offer has been declined, holders of valid patents sue and holders of invalid patents remain inactive.[14] (iii) Denote good quality A's settlement offer as S_I^+ and bad quality A's offer as S_I^-, with $S_I \geq 0$. We know that good quality patentees will demand no less than their certain gains from litigation, so that $S_I^+ \geq (\upsilon - \kappa) - \upsilon_A^D$. Now assume that $S_I^+ > S_I^-$ and B accepts S_I^+ with positive probability. Then bad quality patentees will want to mimic and deviate to S_I^+. The same holds for the opposite case, therefore $S_I^+ = S_I^-$ in equilibrium. As the minimal settlement offer of good inventors is larger than the expected court outcome for B, i.e.

[14] Clearly, there is a continuum of equilibria. In the following, as A can make a take it or leave it offer, we select the equilibrium most profitable to A whenever we consider payoffs from the subgames.

$S_I = (\upsilon - \kappa) - \upsilon_A^D > p_B(\upsilon - \upsilon_A^D + \kappa)$, no feasible settlement offer exists that B is willing to accept.

It is convenient to summarize the results of Lemma 1 and 2 in the table below.

	Outcomes	$\pi_A(\iota \geq \mu)$	$\pi_A(\iota < \mu)$	π_B
Case (C)				
(1) violated	no S, no L	υ	*	0
(1) satisfied	*all* (mixed)	$\upsilon - \frac{\upsilon_B^C - \kappa}{\upsilon + \kappa}\kappa$	$\upsilon - (\upsilon_B^C - \kappa)$	$(1 - p_B)\upsilon_B^C - \kappa$
Case (I)				
(2) violated	no S, no L	υ_A^D	*	υ_B^D
(2) sat, (3) vio	L if valid	$\upsilon - \kappa$	υ_A^D	$\upsilon_B^D - p_B(\upsilon - \upsilon_A^D + \kappa)$
(2) & (3) sat	S	$\upsilon_A^D + p_B(\upsilon - \upsilon_A^D + \kappa)$	*	$\upsilon_B^D - p_B(\upsilon - \upsilon_A^D + \kappa)$

Table 1: Equilibrium outcomes and expected profits of the litigation subgames. (S indicates settlement, L litigation. When holders of invalid and valid patents receive the same profits, these are abbreviated by an asterisk.)

In passing, note that we get the familiar result in the challenge subgame (C) that weaker patents are more likely to result in the settlement outcome. More interestingly, the infringement subgame (I) delivers the opposite result - here, if the perception of patent quality is highest, settlement occurs (both for valid and invalid patents), while intermediate patent quality lets litigation arise (only for valid patents). An analysis focusing on only one type of potential litigation is likely to overlook this fact.

The combination of Lemma 1 and 2 allows us to derive our first proposition:

Proposition 1: *In the Perfect Bayesian Equilibrium of the litigation subgame, B will prefer challenging an existing patent to infringement and inaction iff the following hold: (i) Condition (1) is satisfied, (ii) condition (2) is satisfied and (iii)* $p_B > \frac{\upsilon_B^D - \upsilon_B^C + \kappa}{\upsilon - \upsilon_A^D - \upsilon_B^C + \kappa}$.

The proof of the proposition follows directly from the preceding Lemma 1 and 2. Briefly note that if (1) is violated, B is indifferent between alternative (C) and inaction, and when (2) is violated B always prefers infringement, as this gives him the duopoly payoff. Condition (iii) of the proposition is perhaps the most surprising: This shows that only for patent quality priors *above* a certain threshold B prefers the challenge subgame over infringement. Intuitively, as the patent quality increases, the higher likelihood of having to pay damages makes infringement relatively less attractive.

This lower threshold increases (i.e. challenges become less likely) as the costs of going to court κ increase - this favors infringement, where it is the patent holder who must decide whether or not to initiate court proceedings. It further increases in the difference between duopoly and competitive profits for B, $\upsilon_B^D - \upsilon_B^C$. The higher the protection that B enjoys passively from A's patent, the less likely he is to challenge it, as this would threaten his own cozy situation. On the other hand, if duopoly and free entry profits are relatively similar, say in an industry with other entry barriers, the competitor is more likely to challenge existing patents, unless strategic considerations outside of our model are at play.

Note that (1) gives us a more straightforward upper threshold for patent quality, above which B prefers inaction to the challenging subgame, as his profits from the latter become negative. This threshold is lowered as κ increases - for large court costs therefore the two bounds bypass each other and the set of priors for which challenges may arise becomes empty.[15] If we interpret the prior as a measure of patent quality, we therefore find that for intermediate patent qualities there may be private contribution to the control of patent quality - this result points in the same direction as findings in Chiou (2008), though the mechanism is completely different. Let us next consider the analogue to proposition 1 for the infringement subgame:

Proposition 2: *In the Perfect Bayesian Equilibrium of the litigation subgame, B will prefer infringing an existing patent to challenging it and inaction iff one of the following holds:*

(A) Condition (2) is violated.

(B) (a) Condition (2) is satisfied, (b) $\upsilon_B^D - p_B(\upsilon - \upsilon_A^D + \kappa) > 0$*, and (c)* $p_B < \frac{\upsilon_B^D - \upsilon_B^C + \kappa}{\upsilon - \upsilon_A^D - \upsilon_B^C + \kappa}$

Again, the proof follows directly from lemma 1 and 2. (B) is simply the opposite case from proposition 1, where (b) ensures positive expected profits and (c) lets infringement be more attractive than challenging the existing patent. The case (A) is in some ways more interesting and from the policy perspective also more problematic. In a certain

[15] The set is given by $\frac{\upsilon_B^D - \upsilon_B^C + \kappa}{\upsilon - \upsilon_A^D - \upsilon_B^C + \kappa} < p_B \leq \frac{\upsilon_B^C - \kappa}{\upsilon_B^C}$.

way, this signifies that the patent system as intended has broken down - court costs are so high that going to court is not even feasible for owners of valid patents who will prevail in court with certainty, therefore the competitor can infringe with impunity. At first glance, this may be an artifact of the American rule of assigning legal costs, but also with the British rule the winner of the court battle incurs intangible costs that cannot be recovered, so this case remains relevant.

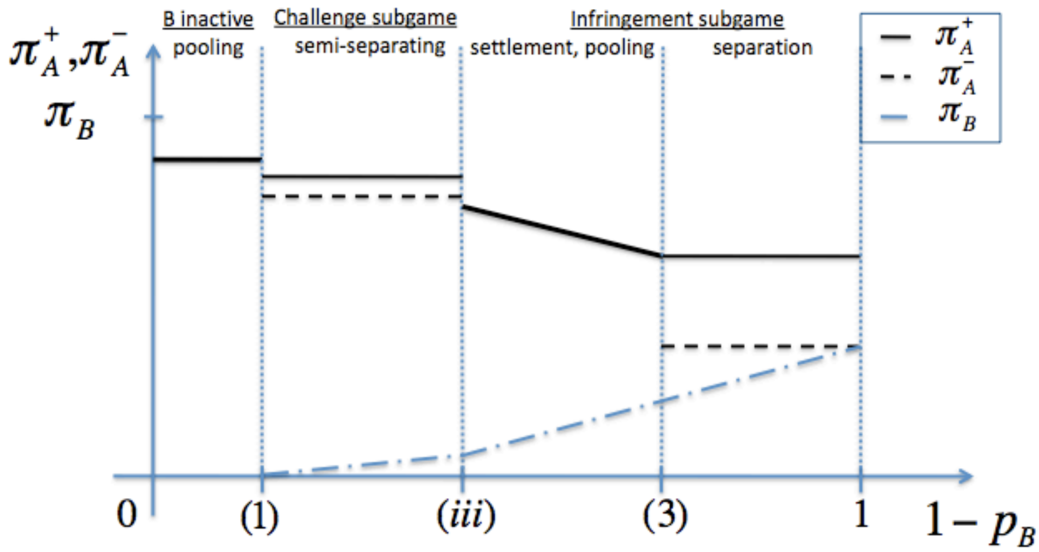

Figure 4: Expected profits of patentees with valid patents (π_A^+), invalid patents (π_A^-) and the competitor (π_B), depending on the prior probability of patent (in)validity.

These two propositions allow us to consider how the profits of the various player types develop as the average quality of a patent decreases given its value and the structure of payoffs. **Figure 4** illustrates the most complex case, given that condition (2) is not violated. As the average quality of the patent decreases (or the share of invalid patents increases), the equilibrium switches from the pure monopoly case to a challenging equilibrium (when condition (1) is satisfied), then to the infringement with settlement (when condition (iii) from proposition 1 is exactly satisfied), to the separating infringement (when condition (3) is satisfied). As discussed above, the area between (1) and (iii) does not necessarily exist, as infringement can dominate challenging from the view of the competitor completely.

Note that there is a wedge between the expected profits of holders of valid and invalid cases only in the sections right of (3) and between (1) and (iii). Overall, the profits of the holders of valid patents are weakly decreasing in the share of invalid patents - note that for certain parameter constellations the value of invalid patents may actually jump upward once, when the regime switches from challenging to infringement, as they can profit from being pooled together with the valid patents.[16] In the following, we will disregard this case in order to avoid having to deal with multiple equilibria and focus on $\frac{\partial \pi_A^-}{\partial p_B} \geq 0$. The following results can also be derived without this assumption, for this it is only necessary that the profits deteriorate if the average patent quality decreases further beyond this point, but proofs become unnecessarily more complicated.

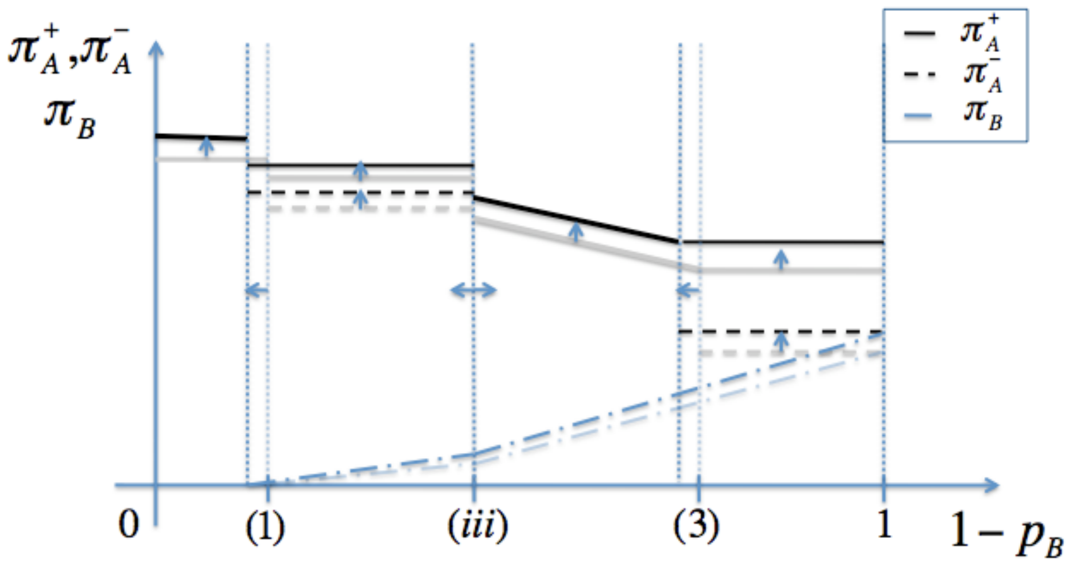

Figure 5: Changes in the payoff structure of the litigation game given an increase in the private value of the patent v.

Finally, Figure 5 displays the effects of an increase in the private value of a given patent v on the profits of the various players. There are two effects: First, profits increase within each outcome for all players. Second, the boundaries of the outcomes shift, most notably, (1) is relaxed so that the threshold for challenging a patent decreases. The other clear effect is that (3) shifts downwards, so that fewer cases reach settlement given

[16]This happens whenever the sum of competitive and duopoly profits more than exceeds the monopoly profits, to be precise, whenever $v < v_A^D + v_B^D + v_B^C \frac{v_B^D - v_B^C + \kappa}{v - v_A^D - v_B^C + \kappa}$.

infringement. The effect on the boundary between challenging and infringing patents is unclear given the assumptions made and depends on the signs and relative magnitudes of the changes in $\upsilon_B^D - \upsilon_B^C$ and $\upsilon - \upsilon_A^D - \upsilon_B^C$. We can note as a result, that as the value of a patent increases, patentees are less likely to enjoy the full monopoly benefits therefrom. Further, if patentees suffer losses from an increase in υ, this is either due to the fact that previously unchallenged patents are then challenged or because previously challenged ideas then suffer from infringement. Most importantly, though, the expected profits of the holders of valid patents are bounded from below at $\upsilon - \kappa$ (or υ_A^D, if (2) is violated) and the expected profits of holders of invalid patents are bounded from below at υ_A^D. These bounds are strictly increasing in υ.

4.1.2 Incentives for Patenting and Idea Generation

Let $\pi_A(p_B, \iota, \upsilon)$ denote the expected outcome from the litigation subgame for A for a given idea and given prior beliefs regarding the patent quality if granted. Obviously, A will patent this idea only if the following condition holds:

$$\phi(\iota)\pi_A(p_B, \iota, \upsilon) + \delta(\Phi)\upsilon \geq \tau_1 \tag{4}$$

A benefits twofold from a patent application. First, there are the expected profits from obtaining the patent, taking the probability of doing so into account if the idea is not objectively patentable. Second, A receives the patent pending status to his idea in any case, which is potentially valuable. If the sum of these two factors exceeds the costs of the patent application, A will decide to patent. Note that there are two potential sources of difference between the expected gains from patenting objectively patentable and non-patentable ideas: $\phi(\iota < \mu) \leq 1$, and $\pi_A^- \leq \pi_A^+$. That is, the latter are less likely to receive a patent and, if they do, the benefits derived from patenting are weakly smaller than those of the former. These considerations lead to the next lemma:

Lemma 3: *For a given structure of market payoffs and level of υ, either no idea is patented, every idea is patented, or a cutoff level $\hat{\iota}(\upsilon)$ exists with the following property: For each idea, if $\iota \geq \hat{\iota}$, A applies for a patent and refrains from doing so otherwise.*

Proof: Rearrange (4) to $\phi(\iota)\pi_A(p_B, \iota, \upsilon) \geq \tau_1 - \delta(\Phi)\upsilon$. Note that the right hand side of the inequality is a scalar for given values of υ. Denote the share of valid ideas among all ideas for a given υ as p_{min}. Consider the lefthand side of the inequality. As $\pi_A^+ \geq \pi_A^-$

and it is non-decreasing in p_B, obviously if $\pi_A^+(p_B = 1, \upsilon) < \tau_1 - \delta(\Phi)\upsilon$, no idea will be patented. Analogously, if $\phi(\iota = 0)\pi_A^-(p_{min}, \upsilon) \geq \tau_1 - \delta(\Phi)\upsilon$, every idea will be patented. Bayesian perfect equilibrium requires that the original prior probability of facing a valid patent $p_B(\Phi, \upsilon)$ be determined by the actual shares in the patenting decision, i.e. the following condition must hold in equilibrium: $p_B(\Phi, \upsilon) = \frac{\int_\mu^1 f(\iota)d\iota}{\int_{\hat{\iota}}^1 \phi(\iota)f(\iota)d\iota} \; \forall \upsilon$. Note that thereby the lefthand side of the inequality is strictly increasing in ι and $\hat{\iota}$, respectively. Therefore, if $\phi(\iota_1)\pi_A(p_B, \iota_1, \upsilon) < \tau_1 - \delta(\Phi)\upsilon$ and $\iota_2 < \iota_1$, then $\phi(\iota_2)\pi_A(p_B, \iota_1, \upsilon) < \tau_1 - \delta(\Phi)\upsilon$.

Logically the following four cases may arise: **(i)** it is not even profitable to patent valid ideas, **(ii)** it is even profitable to patent the least patentable ideas, **(iii)** it is profitable to patent invalid ideas, as long as their quality is above a certain threshold as well as valid ideas and **(iv)** it is *only* profitable to patent objectively patentable ideas.

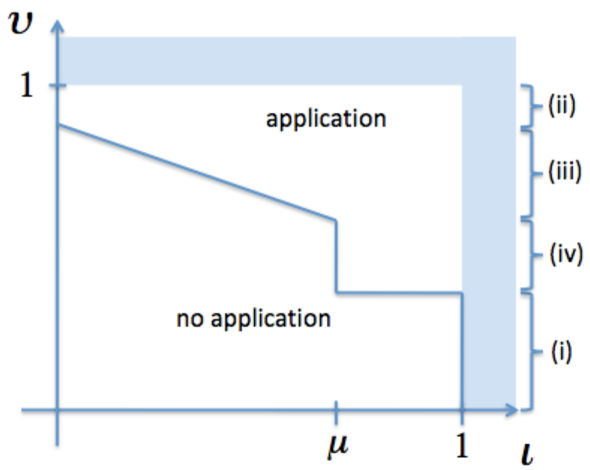

Figure 6: Illustration of Lemma 3 depicting an example in which all 4 cases occur.

Figure 6 illustrates lemma 3 and depicts an example for how patenting decisions are made in (ι, υ) space. It is worth while to consider which of the depicted cases are generic and which only arise under certain circumstances. First note that for every patent application fee $\tau_1 > 0$, case (i) will arise for low enough values υ of valid ideas, i.e. no idea that yields relatively low private benefits will be patented. Next let us focus on case (iv), in which only valid ideas are conferred to the patent office. This case will arise if $\lim_{\iota \to \mu} \phi(\iota) < 1$, for then there is a wedge between the value of patentable and non-patentable ideas. Note that if the share of invalid patents is "small enough" $\pi_A^+ = \pi_A^- = \upsilon$, only as the share exceeds the threshold given by (1) a difference between the two values arises. Case (ii) occurs if $\phi(\iota = 0)\pi_A^-(p_B = 0, \upsilon) \geq \tau_1 - \delta(\Phi)\upsilon$, therefore it disappears if τ_1 is large enough. This case is problematic from a policy perspective, as this means that the patent office is unable to deter *any* bad applications. It is more likely to occur for higher private values and larger returns from pending patents. (iii) arises in the remainder of constellations. Please note that the frontier in (iii) is generally not linear as depicted.[17] Due to the shifts in the boundaries of the litigation frontiers discussed above, the frontier can even be increasing, but only locally.

Finally to close the model, A generates the $(n+1)$th idea as long as the expected profit is larger than its costs, i.e. the following condition holds:

$$\int_0^1 \int_{\hat{\iota}(\upsilon)}^1 [\phi(\iota)\pi_A(\upsilon, \iota) + \delta(\Phi)\upsilon - \tau_1] f(\iota) g(\upsilon)\, d\iota\, d\upsilon \geq C(n+1) - C(n) \tag{5}$$

Clearly, this creates a relatively complex system. For our general setting in which the underlying distribution functions remain not exactly specified a closed form solution to this problem does not exist. This would be a serious shortcoming if the goal of this article were to describe an *optimal* patent system. As our goal is to generate insights concerning the piecemeal reform and adaptation of existing systems, we will be able to use this specification to point out trade-offs and problems that have been mainly overlooked so far in the following section.

4.2 Comparative Statics - Potential Approaches to Reform

4.2.1 Application Fees

One obvious difference between the European and the US patent regime is the significant difference in fees that are imposed for patent application and maintenance. Various authors, most recently Bessen and Meurer (2008a) and Chiou (2008) have suggested to increase patenting fees in order to deter lower quality inventors and raise the average quality of patent applications.

Our setup allows us to dissect the results of such a reform in an extremely straightforward and simple manner. Let us consider a change from τ_1 to τ_1'. This affects the

[17] Instead this resembles a case in which $F(\iota)$ is independent of υ and the expected profits from receiving a patent are either constant or a linear function of υ. This occurs for example if the same litigation-equilibrium arises independent of υ.

patenting decision at two margins. As shown above, the lefthand side of the condition that implicitly defines $\hat{\iota}$, i.e. $\phi(\iota)\pi_A(p_B, \iota, \upsilon) + \delta(\Phi)\upsilon \geq \tau_1$, is strictly increasing in ι given that $\iota < \mu$ for each value of υ. Therefore if the righthand side is increased, the cutoff value $\hat{\iota}(\upsilon)$ must increase as well, which resembles a shift of the frontier in Figure 6 to the right. We will denote the new cutoff value as $\hat{\iota}(\upsilon)'$. Intuitively, this is the desired effect, as fewer invalid ideas are submitted to the PTO office for scrutiny.

There is a very simple second effect though, one that directly affects patentable ideas only. Consider the case in which only valid ideas are patented, therefore $p_B = 1$ and $\pi_A^+ = \upsilon$. Here the patenting decision is defined by $(1+\delta(\phi))\upsilon \geq \tau_1$, which directly implies the minimal private value of an idea that is patented as $\underline{\upsilon} = \frac{\tau_1}{1+\delta(\Phi)}$. Clearly an increase in the application fee τ_1 moves this frontier upward, the effect at the second margin. This second effect is very much undesirable, as it precludes additional objectively patentable ideas from being patented, as the costs outweigh the private benefits.

Further, these effects lead to a strict decrease in the expected profits to be derived from generating new ideas. As a short sidenote: If the smaller number of applications leads to a decrease in the delay between application and patenting δ,[18] then both of these effects increase in magnitude. We summarize these considerations in the following proposition:

Proposition 3: *An increase in the patent application fee from τ_1 to τ_1' leads to an increase in the relative share of valid applications only if the following holds:*

$$\frac{\int_{\underline{\upsilon}'}^1 \int_{\mu}^1 f(\iota)g(\upsilon)\,d\iota\,d\upsilon}{\int_{\underline{\upsilon}'}^1 \int_{\hat{\iota}'}^1 f(\iota)g(\upsilon)\,d\iota\,d\upsilon} - \frac{\int_{\underline{\upsilon}}^1 \int_{\mu}^1 f(\iota)g(\upsilon)\,d\iota\,d\upsilon}{\int_{\underline{\upsilon}}^1 \int_{\hat{\iota}}^1 f(\iota)g(\upsilon)\,d\iota\,d\upsilon} > 0 \tag{6}$$

Proof: The proposition follows directly from lemma 3 and the definition of $\underline{\upsilon}$.

When the condition in the proposition does not hold, the share of valid patents actually *decreases* with a change of the application fee. Considering the marginal effect is useful to better understand this condition. From our previous discussion it is clear that a change in τ_1 will affect the mass of valid as well as the total mass of applications. Let us denote the original share of valid patent application as $\gamma(\tau_1)$ and the total mass of applications as $\alpha(\tau_1)$. Then simple calculus allows us to derive the following much more interpretable corollary to proposition 3:

[18] Arguably, this should not happen in the short run, though, due to the "stockpile" of applications that has accumulated over the years and still has to be dealt with.

Corollary to Proposition 3: *The marginal effect of a change of the application fee τ_1 to τ_1' on the relative share of valid applications will be positive only if the following holds:*

$$|\alpha'| > \frac{(1 - F(\mu))g(\underline{\upsilon})}{1 + \delta(\Phi)}(\gamma(\tau_1))^{-1} \tag{7}$$

From this one can easily determine that the average quality of patent applications is more likely to decrease for increases in the application fee, if the density of patents with the marginal private value $\underline{\upsilon}$ is relatively high. This stems from the fact that on the lower margin only valid ideas are patented. This effect mechanically becomes stronger, the lower the original patent quality is, as the decrease in valid patents then has a stronger weight, as well as for laxer patent regimes (lower levels of μ).

Intuitively, there is a kind of "positive selection" behind this. As the private value of patents decreases, the quality cutoff level for which patents are profitable increases, until at the margin only the best patents survive - if an inventor expects a patent to be of relatively low private value to him, he must be relatively (or completely) sure of its patentability to be willing to incur the costs of patenting. From a policy perspective, these considerations may curb the enthusiasm for raising patent fees somewhat.

4.2.2 Diligence of the Patent Offices

The effects of an increase in the diligence of the PTO appear obvious at first sight. An increase in the inspection intensity should lead to more bad applications being rejected, so that it becomes less attractive to try to pass of bad ideas as patentable. Consider an adjustment of the patent offices policy such that Φ changes to Φ' and thereby $\phi(\iota)' < \phi(\iota)\ \forall \iota < \mu$. By itself, this effect would be purely beneficial. But as assumed above, this improvement comes at a cost, as the returns from pending patents increase to $\delta(\Phi') > \delta(\Phi)$. Note that an increase of the required diligence in the PTO's assessments would in reality not only affect new patent applications but also the existing backlog of past applications that has accumulated.

Revisiting condition (4) yields the following adjusted condition regarding the decision whether or not to

$$\phi(\iota)'\pi_A(p_B(\Phi'), \iota, \upsilon) + \delta(\Phi')\upsilon \geq \tau_1 \tag{8}$$

The effects of such a reform are far less clear cut than we would have expected. Note first that in any case the cutoff level $\underline{v}$ decreases, as the return from valid patents is increasing in δ, which may be considered a positive side-effect. The overall outcome on the other margin depend on the relative sizes in the changes in patenting probability and profits received during the pending phase. Instead of over-stretching our model at this point with regards to the assumptions we have made (for example the linearity of pending-revenues in the private value of an idea), we would like to point out that the profits during the pending phase have a fundamental effect on the current patenting system, yet there are next to no empirical findings with regard to this topic.

4.2.3 Litigation Costs

Various developments in the US and Europe have affected (or will affect) the way patent litigation functions. In the US, the evolution of a specialized appellate court has reduced legal uncertainty (which can be interpreted as a form of cost in our simplified model world) and possibly led to or at least encouraged an increase in patent litigation.[19] Concurrently, in Europe the European Patent Litigation Agreement has been seeking to establish a unified European Patent Court with the expressed goal of reducing patent litigation costs.[20] Our model setup allows us to dissect the effects of such a decrease in some detail.

Let us first consider the changes in the litigation subgame. We collect our qualitative results in Figure 7. Again we observe two kinds of shifts, in the boundaries between the equilibrium regimes and in the payoffs of the individual players. Regarding the boundaries, there is the well known effect that a reduction of court fees leads to less settlements in the infringement game. In our figure this is represented by a shift of boundary (3) to the left. Further, more ideas are going to be challenged at both margins. On the one hand, the competitor will be willing to enter the challenging regime vs. staying inactive earlier on, and on the other hand, challenging becomes relatively more attractive relative to infringement.

The payoffs of the individual players develop in an interesting pattern. First note that the competitor is weakly better off as the court fees decrease - the threat of litigation

[19] See Cook (2007) for more details. Further see Bessen and Meurer (2008a) for a very critical assessment of the recent developments of the patent judiciary in the US.

[20] See for example http://www.epo.org/patents/law/legislative-initiatives/epla.html for details.

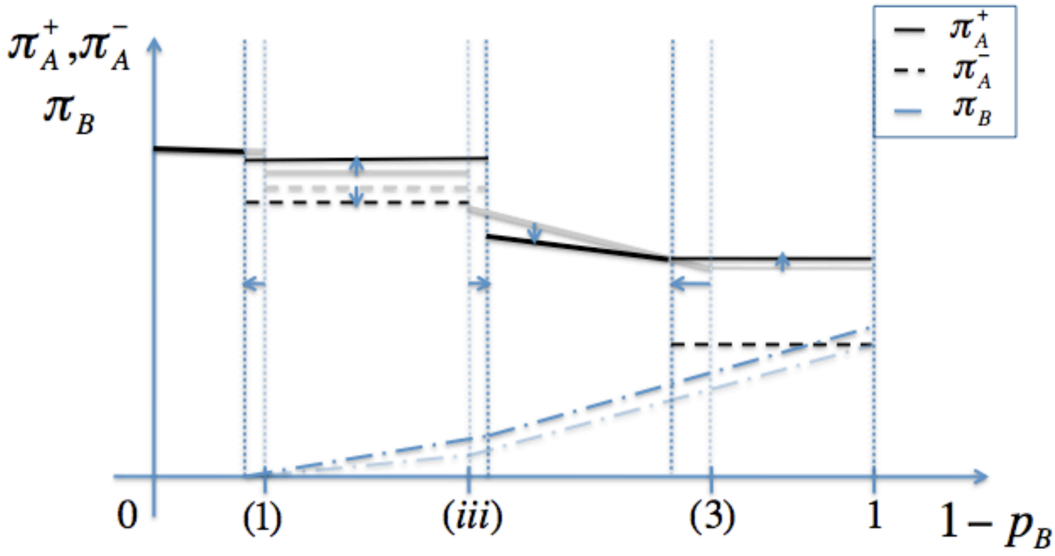

Figure 7: Changes in the payoff structure of the litigation game given a decrease in the costs of going to court κ

bears slightly less weight in the infringement case (which also makes settlement less likely) and the payoff of a successful challenge is strictly higher. The change in the payoffs of a holder of a valid patent are neither strictly positive nor negative in all cases. Given that his patent is challenged, the lower court fees make him better off, yet he is challenged in more cases. Further, the settlement payoffs in the infringement case decrease, which makes patent litigation more likely to arise in this case - note that this may be rather desirable from a policy perspective, as there is a fine line between patent settlements and antitrust issues, see Shapiro (2003). The payoffs of the holders of invalid patents also change in a fashion that may be viewed favorably: The areas, in which a wedge arises between holders of good and bad patents become strictly larger, which makes the efforts of the PTO more likely to bear fruit as discussed above. Except for the cases that proceeded to the challenging subgame prior to the change already, furthermore the profits of holders of invalid patents are weakly lower given a reduction in the court fees.

5 Discussion and Outlook

The current regime of issuing and enforcing patents and the resulting rights of patent owners has faced increasingly harsh criticism in the recent past. There have been calls for completely abolishing the system or at least carrying out a fundamental reform thereof. Due to the enormous investments of the holders of existing patents, it appears relatively unlikely that such a fundamental reform is going to happen soon.

What appears far more likely, is a continuation of the current praxis of little piecemeal improvements (or rather changes). This article tried to add to the understanding of the effects of three of the most considered (or currently implemented) steps that are being taken: An adjustment of the application fees, a policy of stricter scrutiny by the PTO and finally simplified and cheaper adjudication of patent conflicts. Our approach is new in two central regards - first, we take three major aspects of the current system into account in our theoretical model, i.e. policy setting, the role of the PTO, and the interplay of patentees and potential challengers/infringers in the market and in the courts. Second, we take into account that ideas come in a plethora of form and shape and that they may vary continuously in their private value to the patentee and in their patentability (which to a certain extent reflects their value to the community).

From our approach, we gained both new insights and new questions. Allowing the competitor to both challenge or infringe upon an existing patent, we found that below a certain threshold of patent quality, the competitor loses her interest in policing patent quality through challenges but will prefer to infringe upon the rights of the patentee instead. This effect is strongly exacerbated in industries in which there are relatively few barriers to entry - a patent does not only protect its holder. In this context, there is clearly a thin line between tolerating an infringer and anticompetitive settlements of weak patents - focusing on explicit agreements from an antitrust perspective does not do justice to this issue.

Allowing for continuous types allowed us to derive a surprising result with regard to an often contemplated reform step: Raising application fees may actually decrease the average quality of patent applications. This will especially be the case if the distribution of private patent values is skewed towards lower values - which appears to reflect reality very well - and if the standards for patenting are relatively lax. Further, we showed that a decrease in litigation costs through specialized courts (or a centralized court in the case of the EU) may actually make the work of the patent office easier, by forcing a wedge between the payoffs from good and bad patents (or increasing it). Further it can destabilize the infringement "agreements".

Finally, we introduced a problem into our setting, which is so far very little understood - patents yield benefits to their owner before they are approved by the PTO - prior to the decision, even applications that are later rejected may be of considerable monetary or strategic value to their owners. We strongly appeal to empirical economists working in this field to help better understand the issues connected to the "patent pending" stamp. The second empirical issue raised is the interplay between market structure (e.g. barriers to entry) and the outcomes of patent disputes. Our model predicts that the openness of an industry should be negatively related to the probability of a patent being challenged.

The approach of this article was to make relatively few assumptions and see which relatively general results could be derived therefrom - clearly, one logical next step would be to breathe more life into the market setting by imposing tested forms of competition that are well understood together with specific distributions in order to be able to work with closed form solutions. We believe that this more flexible setting is a valuable complement to the existing literature, as it enabled us to point out a number of tradeoffs that are worth studying in more detail.

References

Anton, J. J., and D. A. Yao (2007): "Finding "Lost" Profits: An Equilibrium Analysis of Patent Infringement Damages," *The Journal of Law, Economics & Organization*, 23, 1, 186–207.

Bessen, J. E., and M. J. Meurer (2006): "Patent Litigation with Endogenous Disputes," *The American Economic Review*, 96, 77–81.

——— (2008a): *Patent Failure*. Princeton University Press.

——— (2008b): "The Private Costs of Patent Litigation," *mimeo*.

Boldrin, M., and D. K. Levine (2002): "The Case against Intellectual Property," *American Economic Review*, 92, 209–212.

——— (2008): *Against Intellectual Monopoly*. Cambridge University Press.

Caillaud, B., and A. Duchene (2006): "Patent Office in Innovation Policy: Nobody's Perfect," *mimeo*.

Chiou, J.-Y. (2008): "The Patent Quality Control Process: Can We Afford An (Rationally) Ignorant Patent Office?," *mimeo*.

Choi, J. P. (2006): "How Reasonable is the 'Reasonable' Royalty Rate? Damage Rules and Probabilistic Intellectual Property Rights," *CESifo Working Paper No. 1778*.

Cook, J. P. (2007): "On Understanding the Increase in U.S. Patent Litigation," *American Law and Economics Review*, 9(1), 48–71.

Crampes, C., and C. Langinier (2002): "Litigation and Settlement in Patent Infringement Cases," *The RAND Journal of Economics*, 33(2), 258–74.

Farrell, J., and C. Shapiro (2008): "How Strong are Weak Patents?," *American Economic Review*, 98, 1347–1369.

Hall, B. H., and D. Harhoff (2004): "Post-Grant Reviews in the U.S. Patent System - Design Choices and Expected Impact," *Berkeley Technology Law Journal*, 19(3), 989–1015.

Häussler, C., D. Harhoff, and E. Müller (2009): "To Be Financed or Not... - The Role of Patents for Venture Capital Financing," *CEPR Discussion Paper*, (7115).

Lemley, M. A. (2001): "Rational Ignorance at the Patent Office," *Northwestern University Law Review*, 95, 1495–1529.

Lemley, M. A., and C. Shapiro (2005): "Probabilistic Patents," *The Journal of Economic Perspectives*, 19, 75–98.

Meurer, M. J. (1989): "The Settlement of Patent Litigation," *The RAND Joural of Economics*, 20(1), 77–91.

Prady, D. (2008): ""Is that a Gun in Your Pocket... ?" Advertising and Evaluation," *mimeo*.

Reitzig, M., J. Henkel, and C. Heath (2007): "On sharks, trolls, and their patent prey - Unrealistic damage awards and firms' strategies of "being infringed"," *Research Policy*, 36, 134–54.

Schankerman, M., and S. Scotchmer (2001): "Damages and Injunctions in Protecting Intellectual Property," *The RAND Journal of Economics*, 32(1), 119–220.

Schuett, F. (2009): "Inventors and Impostors: An Economic Analysis of Patent Examination," *mimeo*.

Shapiro, C. (2003): "Antitrust Limits to Patent Settlements," *The RAND Journal of Economics*, 34, 391–411.

Straus, J., and N.-S. Klunker (2007): "Harmonisation of International Patent Law," *International Review of Intellectual Property and Competition Law*, 8, 907–36.