COPYRIGHT NOTICE:

James Bessen & Michael J. Meurer: Patent Failure

is published by Princeton University Press and copyrighted, © 2008, by Princeton University Press. All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher, except for reading and browsing via the World Wide Web. Users are not permitted to mount this file on any network servers.

Follow links for Class Use and other Permissions. For more information send email to: permissions@pupress.princeton.edu

1 The Argument in Brief

Zimbabwe, a country once considered the "breadbasket of Africa," now suffers widespread starvation. Much of this decline can be attributed to the tyrannical policies of President Robert Mugabe—in particular, his disregard for property. In 2000, Mugabe's followers seized land on over one thousand farms owned by white farmers. But when Zimbabwe's Supreme Court ordered the squatters evicted, Mugabe forced the chief justice to resign and physically threatened the remaining justices, who relented. Owners abandoned their property, severely disrupting agricultural production, and within a few years Zimbabwe was wracked by famine.[1]

Even with the rule of law, property systems can fail. A successful property system also requires supportive institutions, and the technical details of property law must make sense. Consider one particular country where many property owners had a hard time enforcing their rights and were often forced to resort to expensive litigation. In one notorious case, a property owner had to assert its rights against more than one hundred parties, an ordeal that involved forty-three separate lawsuits. With so many ostensible trespassers, one might assume the property owner's claim was weak, but as the courts found, this was not the case. Only one suit was dismissed on summary judgment, the owner's claims were largely upheld on appeal, and almost all the defendants settled.

This example does not come from a failed state or a "tinhorn dictatorship." The country in question is the United States; the property is U.S. Patent No. 4,528,643, granted in 1985; the owner who initiated the lawsuits was a company called E-Data; and the alleged violators included a roster of technology companies as well as thousands of small businesses and individuals operating e-commerce websites.[2] This failure of property rights cannot be attributed to a breakdown of the rule of law. Rather it was caused by the failure of patent-related institutions and patent law generally to get the details right. This widespread pattern of alleged violation and litigation would surely be unusual in real estate or personal property in the United States.

Such a rickety system of property rights seems unlikely to be an engine of growth. Burdensome means of enforcement lessen the value of property to its owners. Moreover, property disputes impose costs on other parties. Even though few are sympathetic to trespassers, squatters, and others who seek unjust enrichment, there is good reason to worry about costs imposed on innocent violators. In the case above, many of the defendants believed that they were not infringing upon the owner's rights, and they innocently made investments that turned out to be in violation. Those investments were exposed to unnecessary risk because of unclear property boundaries. A defective property system discourages trade and investment not just by property owners, but also by those who inadvertently face the threat of property related lawsuits.

This book considers patents as a form of property right. If patents work as property, they should reward innovators and encourage investment in innovation. Below, we explore how the laws and institutions of property, including patents, succeed and how they fail. The E-Data example suggests that even in an advanced society with well-developed legal institutions and strong respect for the law, property can fail. Yet this one example might not be entirely representative. We need to go further and ask whether patents work well as property overall.

This question is important because innovators have grown frustrated with the failings of the American patent system. Over the past several years, in newspaper articles and at hearings held by the Federal Trade Commission (FTC) and the National Academy of Sciences (NAS), industry executives have complained in growing numbers that the patent system is broken. In response, Congress has held its own hearings and debated reform. Critics

argue that changes in patent law have created "a legal frenzy that's diverting scientists from doing science."[3] Some even believe that the patent system should be abolished.[4] Others say that the patent system can be fixed with some modest reforms.[5] Still others maintain that the patent system is not broken at all, and that current efforts to reform it are just an attempt to weaken the rights of small inventors.[6]

It is hard to tell who is right, however, because most evidence offered in support of these positions is anecdote, if not myth:

- Defenders of the current system tell stories about the role of patents in protecting small inventors from rapacious corporate giants. But most patents and most litigation do not come from independent inventors, so it is not clear how representative these stories are, or how important small inventors are overall.
- Critics of the system cite patents on a peanut-butter-and-jelly sandwich (U.S. Patent No. 6,004,596), a method of using a backyard swing (U.S. Patent No. 6,368,227), and a method of combing hair over a bald spot (U.S. Patent No. 4,022,227) as evidence of poor patent "quality." Standing alone, however, these patents are not evidence that anything is seriously wrong. Silly patents and patents on unworkable inventions, such as perpetual-motion machines, have been around for at least two hundred years.[7]
- Critics also raise the issue of lawsuits initiated by "patent trolls"—people who obtain broad patents not for purposes of innovation, but solely to ensnare real innovators who might inadvertently cross the boundaries of the trolls' patent. The label "troll" is potent rhetoric, but only a small percentage of patent litigation can be attributed to the most egregious trolls.

Stories about garage inventors are inspiring, while stories about frivolous patents and frivolous lawsuits are troubling, but better evidence is needed to guide patent reform. Without this evidence, it is hardly surprising that some reform proposals seem to be ad hoc.

This book moves beyond anecdote to provide the first comprehensive empirical evaluation of the patent system's performance. We measure

patents against a simple, well-defined yardstick inspired by economic analysis of property rights. Our yardstick weighs the benefit of patents to an innovator against their cost, including the risk of inadvertent infringement. If the estimated costs of the patent system to an innovator exceed the estimated benefits, then patents fail as property.

Some readers might immediately find our objective to be somewhat oddly stated or, perhaps, overreaching. The key limiting qualifier here—the limitation that makes the empirical exercise feasible—is "as property." At the risk of getting a bit pedantic, this phrase requires more careful discussion.

Many readers might think the phrase is redundant in the present context. Some are likely to assume that patents *are* property. What, then, could it mean to ask if they work "as property"? For example, in a paper comparing different ways of providing incentives to innovate, theoretical economists Gallini and Scotchmer (2004) tell us that patents are "intellectual property," which they define as "an exclusive right to market an invention for a fixed time period." As such, it might seem sensible to call patents "property" because exclusion is a hallmark of property. Property rights in land give a farmer exclusive rights to grow crops and bring them to market. If patents provide exclusive rights to market inventions, how could they not work "as property"?

It is important, however, to distinguish idealized depictions of patents from the actual workings of the patent system. Patents do not actually provide an affirmative right to market an invention; they provide only a right to exclude *others* from doing so. This might seem an inconsequential difference, but it has practical significance: other patent holders can block even a patented invention from coming to market. The power to block innovation is especially troublesome when property boundaries are not well identified.

Some of the troubling issues raised by the E-Data patent are explained by this difference. E-Data's patent was for a kiosk that produced digital audio tapes and the like in retail stores, but they interpreted this patent to cover a very broad swath of e-commerce. On the other hand, IBM holds hundreds of patents related to e-commerce, but this did not prevent E-Data from threatening to block IBM from marketing its own innovative

e-commerce products. To market its own patented technologies, IBM was forced to pay E-Data for a license. Similarly, Research in Motion (RIM) holds patents on its popular BlackBerry personal-communication device, but this did not prevent NTP, a patent licensing company, from famously suing RIM for patent infringement. To avoid being excluded from the market for its own patented invention, RIM paid NTP more than a half-billion dollars.

Gallini and Scotchmer present idealized notions of patents and property that might be useful for some theoretical inquiries.[8] Empirical investigation, however, requires us to be mindful of the ways in which real patents might fall short of such stylized concepts of property. If examples like E-Data and RIM are typical—and this is the kind of empirical question we explore in depth below—then patents will perform quite differently from the property ideal. Below, we estimate just how far actual patents fall short of that ideal. Patents work well as property for some kinds of technology and given the right institutional setting. Patents fail as property for other kinds of technology and given the wrong institutional setting.

Overall, the performance of the patent system has rapidly deteriorated in recent years. By the late 1990s, the costs that patents imposed on public firms outweighed the benefits. This provides clear empirical evidence that the patent system is broken. Both our empirical analysis and our comparative institutional analysis provide clues about the causes of this deterioration—and about what might be done to fix it.

Our focus is on the American patent system, and some of the problems we identify with the American patent system are unique. Nevertheless, there are two reasons that our analysis has relevance to innovation in other countries. First, many other patent systems are under some pressure to become more like the American patent system. For example, Japan and Europe have loosened restrictions against software patents. Second, patent rights and patent litigation are global matters. Important inventions are usually patented in all major markets. This means that patent holders can choose where to litigate. Increasingly, patent disputes are being litigated in the United States, usually resulting in worldwide settlement agreements. Indeed, European inventors file more lawsuits in the United States than they file in any European country other than Germany. This means that

the United States patent system directly affects firms and innovators in Europe, Japan, and elsewhere.

Patents as Property

We begin by comparing patents to tangible property. Lawyers and legal scholars—perhaps because they have endured at least a semester of training in property law and are therefore aware that things might not be so neat—tend to speak of patents not as a form of property, but as *analogous to* other forms of property. Some argue that the analogy might not be appropriate (Lemley 2005), others that the analogy is long-standing (Mosoff 2007), but most recognize that the law and institutions of property systems are complicated and patent law necessarily diverges from the law of tangible property.

We begin our inquiry in chapter 2 by looking at the appropriateness of this analogy, comparing the property-like features of patent law to features of the law of tangible property. We have already noted one important difference: patents do not provide an affirmative right to use an invention. More than one person can use an invention at a time and more than one inventor can claim rights over an invention. Many people can even invent the same technology independently at the same time. In contrast, tangible property is a "rival" good—that is, only one person can use it at a time. This means that the *right to exclude* others more or less conveys an affirmative *right to use* tangible property. As we shall see, this difference between inventions and tangible property is important.

In many other ways, however, patent law shares essential doctrinal features with the law of tangible property. Specifically, patents provide partial rights to exclude others from using an invention as well as rights to transfer ownership. Just as property rights provide incentives to invest in the acquisition, development, and maintenance of tangible property, patents potentially provide incentives to conceive a new technology ("invention"), develop it into a commercial product or process ("commercialization"), and put it to use ("innovation"). Such "innovation incentives" are central to the Constitutional mandate to "promote the progress of . . . the Useful

Arts," which the framers set out when empowering Congress to devise a patent system.

But property and patents only *potentially* provide these incentives. Our review finds well-known evidence that property systems sometimes fail to provide such incentives efficiently:

- Property rights can fail when their validity is uncertain. Such was the case when the transition from Mexican to American rule in California during the nineteenth century clouded the validity of land titles granted under Spanish and Mexican rule. This uncertainty led to squatting and a decline in agricultural productivity.
- Property rights can fail when rights are so highly fragmented that the costs of negotiating the rights needed to make an investment become prohibitive. Such was the case with Russian retail establishments in that country's transition to a private economy. Ownership rights to stores were granted to large numbers of parties, making it too difficult for any one group to obtain the required permissions to operate each store. The stores were often shuttered while street vendors conducted a busy trade nearby.
- Property can fail when boundary information is not publicly accessible. In many less-developed nations, cumbersome regulations discourage impoverished people from recording property boundaries. This limits their ability to trade that property or to use it as collateral for obtaining loans.
- Finally, property rights can fail when the boundaries of the rights are not clear and predictable. This problem sometimes arises with property extracted from nature, such as mineral rights. For example, mineral veins beneath the surface of the earth twist and intersect in unpredictable ways. Such a boundary-related failure in the copper mines of Butte, Montana, led to a violent struggle between rival claimants.

These failures emphasize the importance of implementation in property rights systems. The economic effectiveness of any property system depends not just on what it sets out to do, but also on the laws, regulations, institutions, and norms that implement the system. Consequently, the

doctrinal similarity between patent law and the law of tangible property can obscure important differences in economic performance that arise because these doctrines are implemented differently. Patents might not work well as property if patent law is not implemented effectively; the messy details of how patents work matter.

The Notice Function: If You Can't Tell the Boundaries, It Ain't Property

We can identify one very important difference between the way property rights and patent rights are implemented. This difference concerns the "notice function" of property. An efficient property system notifies non-owners of property boundaries. For example, land rights have a well-developed and efficient system to notify third parties of boundaries. Because of this, only rarely does someone invest millions of dollars constructing a building that encroaches on someone else's land without permission. Far more typically, would-be investors "clear" the necessary rights before investing. They locate markers, check land deeds, conduct surveys, and so forth, in order to determine the adjacent boundaries. They then either negotiate rights to the needed land or design the building to avoid encroachment.

The notice function does not always work so well with patents. For example, the E-Data dispute arose because hundreds of parties, including some very large companies, ignored, did not see, or misunderstood the boundaries created by the patent in question. That patent, awarded to Charles Freeny Jr., was entitled a "System for Reproducing Information in Material Objects at a Point of Sale Location." Its unhelpful title obscures the fact that Freeny actually invented a kiosk for producing music tapes or other products in retail stores using digital information. But, as we have seen, the patent was asserted against thousands of companies doing e-commerce, a rather different technology.

Why did notice fail so completely in this case? For one thing, a prospective technology investor needs to check a very large number of patents. According to David M. Martin, CEO of a patent risk management firm, "if you're selling online, at the most recent count there are 4,319 patents you

could be violating. If you also planned to advertise, receive payments for, or plan shipments of your goods, you would need to be concerned with approximately 11,000."[9]

But even if a website developer could check all these patents, it would be very difficult to know what their boundaries are. The boundaries of the E-Data patent depend on the meaning of abstract phrases such as "point of sale location," "information manufacturing machine," and "material object." Consider, for example, the meaning of "point of sale location." This is a bit of computer and retail-industry jargon first used when electronic terminals replaced cash registers. It refers to the location within a store where items are checked out and transactions take place. Did this term in the patent claim limit the patent to transactions in retail stores, or did it cover all e-commerce, including transactions that might take place in buyers' offices or even in their bedrooms? The district court limited the patent right to retail locations. In 2001, however, the Court of Appeals for the Federal Circuit, using legal rules that place little weight on actual industry usage or on dictionary definitions, concluded that the "point of sale location" included bedrooms, offices, and anywhere else with an Internet connection. Thus, sixteen years after the patent was granted, it was given boundaries that many people, including a district court judge, would find surprisingly broad. In the interim, the correct boundaries of this patent were essentially unknown. The patent offered poor notice.

Poor notice causes harm because it subjects technology investors to an unavoidable risk of disputes and litigation. The expected cost of inadvertent infringement imposes a *dis*incentive on technology investors. Potential innovators consider not only the reward that they might reap from owning patents, but also the risk of being sued for infringing upon the patents of others. Clearly, if the risk of inadvertent infringement is too great, the net incentives provided by the patent system will be negative, and patents will fail as a property system. This is similar to the failures that occurred with Mexican land grants in California, with Russian retail store ownership, and with the copper-mine war in Butte, Montana, noted above.

Establishing notice is often inherently easier for tangible property because, as opposed to patents, tangible property is a rival good. This means that active possession of tangible property is often sufficient to inform the world about what is owned and who owns it—consider, for example, the

shirt on one's back. For nonrival inventions, such as RIM's technology, however, the fact that RIM independently developed and actively possesses the technology does not help clarify the relevant patent boundaries and ownership.

In addition, the implementation of patent notice suffers important deficiencies. In chapter 3 we explore several institutional differences between patents and the property system for land that might make the former particularly prone to notice problems. These institutional features affect patent notice and are thus central to our analysis:

1. *Fuzzy or unpredictable boundaries*: Surveying land is inexpensive, and the survey boundaries carry legal weight. While surveyors can plainly map the words in a deed to a physical boundary, it is much harder to map the words in a patent to technologies, as the E-Data patent dispute illustrates. Not only are the words that lawyers use sometimes vague, but the rules for interpreting the words are also sometimes unpredictable. Although innovators can obtain expensive legal opinions about the boundaries of patents, these opinions are unreliable. There is thus no reliable way of determining patent boundaries short of litigation.

2. *Public access to boundary information*: The documents used to determine boundaries for both land and patents are eventually available to the public. It is possible, however, for patent owners to hide the claim language that defines patent boundaries from public view for many years, a practice that is becoming increasingly frequent.

3. *Possession and the scope of rights*: Generally, tangible property rights are linked closely to possession, hence the well-known expression, "possession is nine points of the law." Patent law also requires possession of an invention, but often this requirement is not rigorously enforced. Courts sometimes grant patent owners rights to technology that is new, different, and distant from anything they actually made or possessed. Not surprisingly, this practice makes patent boundaries especially unclear in fast-paced fields such as biotechnology and computer software development. It certainly seems that E-Data was granted ownership of technology that was far removed from what Charles Freeny Jr. actually invented.

4. *The patent flood*: Clearance costs are affected by the number of prospective rights that must be checked for possible infringement.

Investments in land or structures rarely involve many parcels of land, and property law discourages fragmentation of land rights. In contrast, investments in new technology often need to be checked against many patents—even thousands, in the case of e-commerce. Although the patent system has features that discourage patent proliferation—notably the requirement that an invention not be obvious—empirical evidence suggests these are not working well.

These differences mean that patents might diverge significantly from an ideal property system that grants an inventor a well-defined, exclusive right to develop a technology and bring it to market. Because of such differences, patents might not work well as property. Whether or not they do is the empirical question at the heart of this book.

Empirical Evidence: Do Patents Work As Property?

Do patents give inventors positive net incentives to invest in innovation? An answer requires careful attention to the details of the patent system and the markets for innovation. The empirical evidence must account for incentives from many sources, including exclusion of competitors, licensing, and sale of the patent. We must be careful to distinguish between patent based incentives and other incentives to invest in innovation. We must also account for disincentives that arise indirectly from the threat of litigation. We study how the pattern of incentives varies over time, and across industries, technologies, and types of inventor.

Our question is simpler and more basic than the questions economists often ask when evaluating policy. Economists like to ask whether policies increase "net social welfare," a generalized measure of the overall well-being of society. Short of that, economists like to ask whether innovation policies increase innovation or R&D spending. But these are even more difficult and complicated questions to investigate empirically. Many interrelated factors can influence R&D spending, innovation, and the resulting social welfare, so it is difficult to disentangle these to determine the independent influence of patents. Not surprisingly, economic studies that

attempt to answer these more difficult questions typically have arrived at inconclusive results.

Our approach, instead, is to ask a more limited question. We can determine, with reasonable accuracy, whether or not patents provide net positive incentives for a given group of inventors. This does not tell us whether patent policy is optimal or not. To the extent that incentives are positive, we cannot tell whether they are too big or too small relative to the social optimum. There are many factors we cannot measure that go into a calculation of the optimal incentive. On the other hand, if patent incentives are negative, then they fail as property in a basic sense. In this case, patents do not do what they are supposed to do, and it is not likely that they will spur innovation and increase social welfare.[10] Even though society might receive benefits far beyond the innovator's profit, if patents discourage innovators on net, then patents will not help realize these benefits.

We begin our empirical analysis in chapter 4 by reviewing the literature. It has been almost fifty years since the empirical evidence on patents was last comprehensively reviewed. The reviewer, Fritz Machlup (1958), concluded that it was not possible to decide whether patents were good or bad policy instruments. In the interim, a wide variety of research has looked at the performance of patent systems.

We review this scholarship not to determine whether patents are good or bad policy instruments in general—the discussion above suggests that firm conclusions of this sort might be very difficult to reach. Rather, we simply ask whether a nation's patents seem to have a similar effect on its economic performance as do other property rights in that nation. If they do not, this suggests that the implementation differences between patents and other property rights might be significant. Even though economic performance is ultimately influenced by *global* property rights, the contrast between a nation's patents and its other property should reveal important differences or similarities.

Specifically, the research we review includes:

- *Historical research on the Industrial Revolution*. Although property rights and markets fostered economic growth and innovation throughout Europe and the United States, patents played a much more limited role. In Britain, few major inventors received much benefit from patents,

although in the United States more did benefit. More generally, however, countries without patents were just as innovative during this period as those that had patents.

- *Statistical studies that compare the performance of countries over time*. These studies use indices of the strength of property rights or the strength of patent rights to explain each country's economic growth rate. Although general measures of property rights exhibit robust correlations with economic growth, measures of patent rights do not. Patents might still play an indirect role, however. Patent rights appear to be somewhat correlated with R&D spending, although this relationship exists only among more-developed countries, and it is not clear whether patents cause R&D or vice versa.
- *Studies of economic experiments*. These studies explore what happens when legal rights change. Some researchers have explored the role of property rights in the transition of former Soviet-bloc countries to market-based economies. Those countries that developed property institutions to support a robust market economy have experienced strong economic growth after an initial period of sharp decline. This success apparently depends, however, on the establishment of specific supportive institutions, including market-oriented legal systems, commercial banking, regulatory infrastructure, and labor-market regulation. Countries that introduced private property and markets without developing these institutions have experienced persistently declining per capita income. By contrast, economic experiments that extended or strengthened *patent* rights do not seem to show clear evidence of increased innovation, except, perhaps, to a limited degree among the wealthiest nations.
- *Miscellaneous research*. Case studies present a convincing argument that patents are critical for investment in R&D in the pharmaceutical industry. On the other hand, survey evidence suggests that in most other industries, patents do not pose much of a barrier to imitation, and firms rely mainly on other means, such as lead-time advantages and trade secrecy, to obtain returns on their R&D investments. Moreover, several studies suggest that a moderate degree of competition might actually spur innovation.

In summary, patents do not work "just like property." While they do play some role in promoting innovation and economic growth, that role is limited and highly contingent compared to the role property rights normally play in promoting economic growth. The laws and institutions that implement property rights might be harder to get right for patents than for tangible property rights.

Nevertheless, patents might still work effectively, even if they have a more limited impact on economic growth than property rights for other assets. To arrive at a more definitive evaluation, we need to perform a careful accounting of the incentives and disincentives for investing in innovation that patents provide. We do this by drawing on estimates found in the literature, and on some of our own estimates.

Figure 1.1, copied from chapter 6, conveys the basic calculations we make for United States public firms. Because chemical patents, especially on pharmaceuticals, are much more valuable and much less likely to be litigated, we display the chemical and pharmaceutical industries separately from other industries. The heavy, solid lines show the annual aggregate costs to these firms of defending against patent litigation. This estimate includes not only direct legal costs of litigation, but also business costs such as loss of market share or the costs of management distraction. The dashed lines represent an estimate of the incremental annual profit flow from all patents worldwide associated with inventions patented in the United States. We derive these estimates, based on a review of over twenty research papers, in chapters 5 and 6.[11]

The incremental profit flows represent the gross positive incentives provided to innovators by worldwide patents above and beyond the profits that could be earned without patents. Litigation costs represent an important disincentive to innovation. A firm looking to invest in innovation will consider the risk that the innovation will inadvertently expose it to a patent infringement lawsuit. Since infringement lawsuits are usually filed against firms exploiting new technologies, development of a new technology exposes the innovator to risk of inadvertent infringement if patent boundaries are hidden, unclear or unpredictable.[12] That risk weighs against the profits that can be made from innovation. Of course, firms are both patent holders and potential defendants, so a comparison of profit flows and litigation costs for a group of firms should reveal the sign of net incentives.[13]

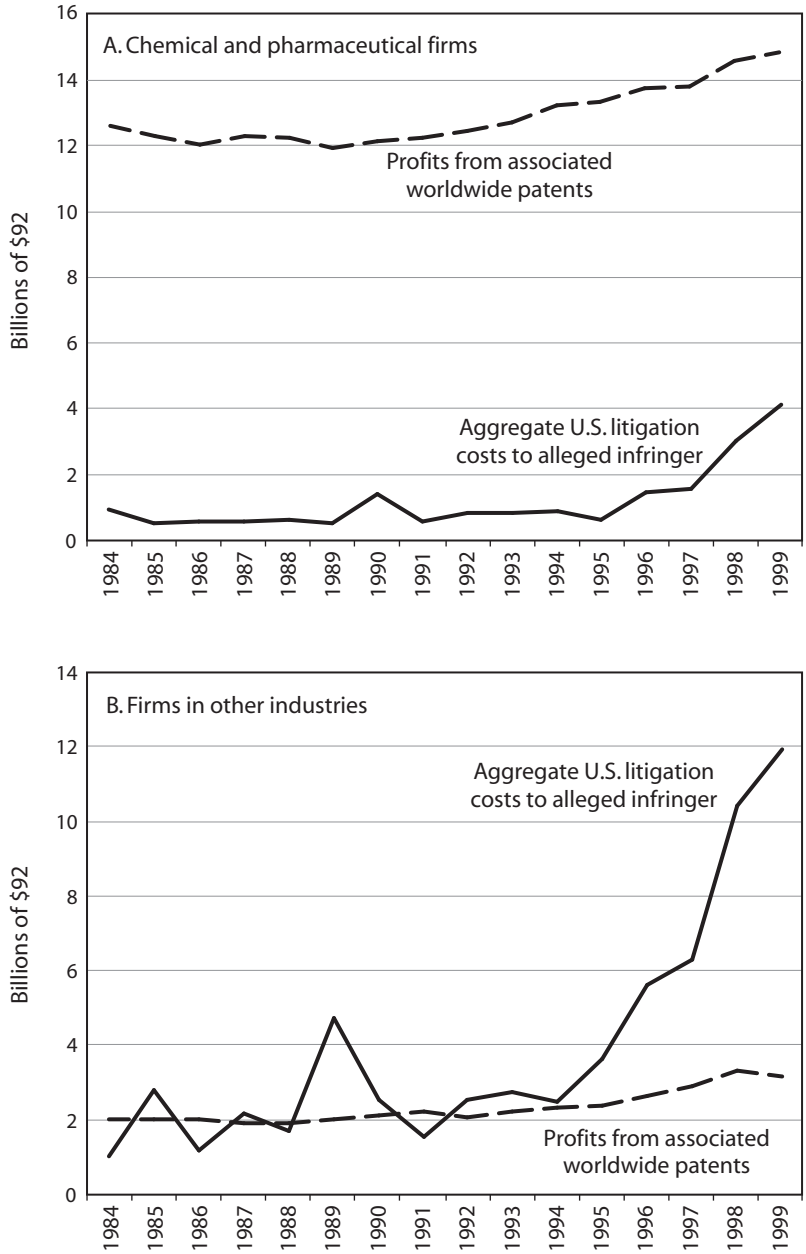

Figure 1.1. Aggregate Profits from Patents and Aggregate Litigation Costs for U.S. Public Firms. *Note*: This is a duplicate of figure 6.5, below.

The results in figure 1.1A show that chemical and pharmaceutical firms earn far more from their patents than they lose to litigation. But for other firms, figure 1.1B tells a simple but dramatic story: during the 1980s, these firms might have, at best, broken even from patents, but in the mid-1990s litigation costs exploded. By almost any interpretation, the United States patent system could not be providing overall positive incentives for these United States public firms by the end of the 1990s. The risk of patent litigation that firms faced in their capacity as technology adopters simply outstripped the profits that they made by virtue of owning patents. A firm looking to invest in an innovative technology during the late 1990s, taking this risk into account, would expect the net impact of patents to reduce the profits from innovation rather than to increase them. Moreover, preliminary data for more recent years suggest that this problem has gotten worse since 1999.

Note that patents *do* provide profits for their owners, so it makes sense for firms to get them. But taking the effect of *other* owners' patents into account, including the risk of litigation, the average public firm outside the chemical and pharmaceutical industries would be better off if patents did not exist.

Moreover, figure 1.1B understates the extent to which costs exceeded benefits for several reasons: disputes settled before a lawsuit was filed are not counted, nor are foreign disputes; this comparison ignores the costs of obtaining patents and clearance; and for a variety of reasons, the estimates of worldwide patent profits are biased upwards, while the estimates of litigation costs are biased downwards.

The patent system clearly provides large positive incentives for innovators in the chemical and pharmaceutical industries. Also, small firms generally receive benefits that exceed costs, but the net incentives for these patentees are not large. We will comment more on these exceptions below.

Interpreting the Results

To understand the meaning of the evidence in figure 1.1, we explore several related issues. First, what is driving this surge in the cost of litigation? In

chapter 7 we look at a variety of alternative explanations. The increase in aggregate litigation cost is mainly driven by the increasing frequency of litigation, which has roughly tripled since the 1980s.[14] Yet when we look in detail at what determines the rate of litigation, we find that only a small part of this increase can be explained by measurable factors, such as how much the parties to a lawsuit spend on R&D or how many patents they have. This suggests that most of the increase arises from unmeasured factors that might include legal, institutional, and technological changes. We explore several possible factors, including deterioration of patent notice, industry-specific factors, greater rewards for litigation, a general increase in litigiousness, the rise of patent "trolls," and the declining quality of patent examination.

We can directly rule out several of these explanations. All industries appear to have experienced a rapid increase in patent litigation, although the increase seems somewhat more rapid in software-related industries. This means that industry-specific factors are unlikely to explain most of what is happening. Also, business-to-business litigation has not been increasing in general, so we cannot attribute the increase in patent litigation to an overall rise in litigiousness. In addition, we find no evidence to suggest that the rewards that patent holders gain from litigation increased in the 1990s, although they might have increased during the 1980s.[15]

We also considered the role of patent "trolls," which we define narrowly as individual inventors who do not commercialize or manufacture their inventions. One story claims that the increasing availability of patent litigators willing to work on contingency fees has spurred lawsuits by such trolls, who might otherwise be unable to afford litigation. The share of lawsuits initiated by public firms has not declined, however, nor has the share of lawsuits involving patents awarded to independent inventors increased. This suggests that the increase in litigation cannot be mainly attributed to patent "trolls," at least through 1999. Of course, if we use a broader definition of "troll" that includes all sorts of patentees who opportunistically take advantage of poor patent notice to assert patents against unsuspecting firms, then troll-like behavior might be a more important explanation. Indeed, if patent notice is poor, then all sorts of patent owners might quite reasonably assert patents more broadly than they deserve. But then it is more appropriate to attribute the surge in litigation to poor patent notice, not to trolls per se.

In fact, the distinctive pattern of litigation over time and across technologies does provide support for an explanation based on the deterioration of patent notice. Several changes to patent notice occurred during the mid-1990s, including the way that courts interpret patent claims, increased "hiding" of patent claims while applications are under review at the Patent Office, problematic legal decisions in software and biotechnology that extended the reach of patents to technologies far beyond what was actually invented, and the growing flood of patents that began during the mid-1980s and gathered strength during the 1990s. Many of these changes are the specific work of the Court of Appeals for the Federal Circuit, the specialized appeals court for patents that was established in 1982. In any case, these changes provide a natural explanation for the concurrent increase in litigation.

In addition, the pattern of litigation costs across technologies is consistent with differences in patent notice. Litigation costs are particularly low for patents on chemical compounds, including pharmaceuticals. At the same time, the value of these patents is much higher than the value of other patents—perhaps, in part, because litigation costs are low and enforcement is effective. Economists have long recognized that patents on chemicals work particularly well because these patents have very well-defined boundaries.[16] In contrast, economists recognize that complex technologies, such as electronics and computers, might have relatively poorly defined patent boundaries. Patents on complex technologies have higher litigation rates and lower values than chemical patents. By the late 1990s, these patents generated more litigation cost than profit.

Software patents, in particular, often have boundaries that are especially difficult to determine, for reasons we explore further in chapter 9. Software patents have even higher litigation rates and a high frequency of appeals over the meaning of patent claims. Not surprisingly, the costs of litigation for software patents far exceed the profits. The distinctive pattern of litigation rates across technologies thus supports the notion that patent notice might explain differences in patent value.

The deterioration of patent notice might also be roughly associated with a decline in patent quality, broadly conceived. Many critics equate low patent quality with frequent issuance of invalid patents. These critics contend that poor examination allows invalid patents to be issued for inventions

that are obvious or lack novelty. Specifically, they assert that inadequate search of previous patents and publications causes examiners to overlook novelty and obviousness problems. Other critics attribute patent quality decline in part to the Federal Circuit's proclivity to weaken the legal test of obviousness. These two sources of patent quality decline contribute to the patent flood and make clearance difficult and costly, leading indirectly to litigation. Our broader conception of patent quality acknowledges problems with novelty and obviousness, but our evidence shows that quality problems are more fundamentally connected to problematic boundaries associated with patents that are vaguely worded, overly abstract, of uncertain scope, or that contain strategically hidden claims.

The narrow conception of patent quality decline does not explain the surge of patent litigation or the pattern of litigation across technologies. Perhaps there has been a recent surge of invalid patents granted, but no such surge appears in the data on litigation outcomes. Similarly, invalidation rates are not higher for technologies featuring higher litigation rates. This suggests that patent examination search quality is not primarily responsible for the increase in costly litigation by itself, although it might well be a contributing factor and it might also be a problem for other reasons.

This analysis leads us to the conclusion that during the late 1990s, the American patent system failed as a system of property rights for public firms. It did so because it failed to provide clear and efficient notice of the boundaries of the rights granted.

Small Inventors

But this evidence of failure applies only to one group of inventors—namely, those at public firms. Now this is a large and very important group of inventors, especially if we assume that the main purpose of patents is to provide incentives to invest in R&D—this group of firms is responsible for about 90 percent of R&D spending. Nevertheless, some important inventions are made by small inventors, including independent individuals and small nonpublic firms. Perhaps the patent system works sufficiently well for these small inventors to offset its other failures. Indeed, some people claim

that almost all "breakthrough" inventions come from small inventors, and their interests should be paramount in debates about patent reform.

Chapter 8 explores several issues regarding small-inventor patents. There are good reasons to think that small inventors make important inventions. This is not true of all types of small inventors, of course; many small inventors patent games, simple machines, and other low-tech inventions. Nevertheless, many small inventors do make important high-tech inventions. But there is no evidence to suggest that *most* breakthrough inventions come from small inventors. What limited evidence exists—for example, the characteristics of inventors nominated to the National Inventors Hall of Fame—suggests that most recent major inventions originated in large organizations, although a significant minority of important inventions are developed by independent inventors or inventors working in small firms.

How does the patent system perform for small inventors? In our analysis of public firms, we find that small public firms enjoy positive incentives from patents—their litigation costs are lower than the profits they receive from patents, although their absolute level of profits from patents is not large. Other small inventors are also likely to enjoy a positive incentive, but we lack the data to estimate their litigation costs. Certainly, the many small inventors who do not commercialize any technology have little to fear as defendants in patent lawsuits. Even so, we find troubling evidence that patent notice adversely affects small inventors, too. Patent notice problems impair the market for technology and rob many small inventors of the larger reward they could get from licensing or selling their patents in a world with good patent notice.

The troubling evidence is this: all types of small inventors, including small firms, realize substantially less value from their patents than do large firms. This is true for the independent inventors who work in low-tech fields, as well as for small public firms in many high-tech industries. Indeed, relative to large firms, many small inventors, even small high-tech firms that go public, forgo patents entirely, relying instead on trade secrecy and other means of protecting their profits deriving from innovations. The patent system works for small inventors, but does so only weakly. Why? In part because small inventors do not have access to the resources needed to commercialize inventions. They cannot quickly ramp up manufacturing

and marketing, they do not have established distribution channels, and they cannot easily finance acquisition of these assets.

Lack of such complementary assets would not be a problem if markets for technology worked better. Small inventors who lack resources should be able to sell or license their technology to large firms who have those resources. Indeed, technology markets are often the best means that small inventors can employ to capture value from their inventions. In a world with competitive and efficient technology markets, licensing royalties and sales contracts would deliver value to small inventors comparable to the value that large firms gain from their own patents. The fact that small inventors actually gain much less from their patents, however, indicates that these markets do not always work very well.

Better patent notice would improve technology markets in two ways. A direct improvement flows from clearer property rights. Unclear property rights increase bargaining costs and the probability of bargaining breakdown. Better patent notice makes technology markets more efficient, and hence more attractive to small inventors.

An indirect and possibly larger benefit flows from the impact of notice on buyers in technology markets. When potential buyers face substantial risk of patent litigation, they cannot profit as much from the technology they seek to exploit and are therefore unwilling to pay as much for the technology. Better notice would reduce the risk of inadvertent infringement and any ensuing litigation, increase the willingness of buyers to pay for technology, and increase the value of patents to small firms who sell in technology markets.

Small inventors and large firms alike suffer from poor patent notice.[17] The positive incentives that small inventors receive from patents give us no reason to be sanguine about the current state of the patent system.

The Particular Problem of Software Patents

As we noted above, the patent system performs particularly poorly for software patents. Software is a vital technology and, as we shall see, software

patents contribute substantially to the overall failure of the patent system for public firms. We explore the reasons for this in chapter 9.

Software patents have been controversial in part because the software-publishing industry grew up largely without patents and most computer professionals oppose patenting software. But judicial decisions during the 1990s eliminated certain obstacles to software patents, and now close to 200,000 software patents have been granted.

Some argue that there is nothing different about patents on software, and if there are any problems, these will be resolved as the Patent Office adapts to this new technology. Some say that because the software-publishing industry remains innovative, patents cannot be hurting innovation. But evidence about the software-publishing industry is not definitive; the majority of software firms still do not obtain patents, and most software patents are awarded to firms in other industries, chiefly the manufacture of computers, semiconductors, and electronics.

Critically, software patents *do* seem to exhibit some marked differences from other patents when it comes to litigation costs. Software patents are more than twice as likely to be litigated as other patents; patents on methods of doing business, which are largely software patents, are nearly *seven* times more likely to be litigated. And, despite being a relatively new area for patenting, software patents accounted for 38 percent of the total cost of patent litigation to public firms during the late 1990s. This does not appear to be a temporary problem that is dissipating as the Patent Office adapts—the probability that a software patent will be litigated has been *increasing* substantially rather than decreasing.

Why are software patents more frequently litigated? In a word, abstraction. In chapter 9, we will elaborate upon what we mean by "abstraction" and how it affects patent notice, but for the present consider that software is an abstract technology, and this sometimes makes it more difficult, if not impossible, to relate the words that describe patent boundaries to actual technologies. In this context, it is helpful to recall the abstract concepts described in the claims of the E-Data patent—"point of sale location," "information manufacturing machine," "material object," and so on. These words reach far beyond the actual kiosk technology that Charles Freeny Jr. of E-Data invented, yet during the course of litigation it was not clear what, precisely, they covered. In other cases, the words in some broad

software patents seem clear enough, but because the patents claim technology far beyond what was actually invented, judges will sometimes interpret the claims narrowly (for example, see the discussion of Wang's Patent No. 4,751,669 in chapter 9). But it is hard to predict which broad claims will be narrowed. This becomes another cause of boundary unpredictability that contributes to inadvertent infringement and, ultimately, to litigation.

Of course, software is not the only technology that can be described in abstract terms. Indeed, the problem of abstraction in patents has been recognized at least since the eighteenth century, when British law attempted to exclude patents on general "principles of manufacture" as opposed to specific inventions. In the United States, judges also developed doctrines to exclude patents with abstract claims during the nineteenth century.

Nevertheless, there are two major reasons why abstraction poses a particular problem for software. First, as we will discuss in chapter 9, the Court of Appeals for the Federal Circuit has tolerated more abstraction in software patents than seems warranted by these patent doctrines. Second, software is inherently more abstract than other technologies. Indeed, it is well known among computer scientists that software technologies (algorithms, system structures) can be represented in many different ways, and it might be difficult to know when alternative representations are equivalent. This means that the technology claimed in a patent can be difficult to distinguish from alternatives; it might be hard to know whether a given patent claims an invention that is different from previous inventions, or whether an allegedly infringing program is different from the claimed technology. If computer scientists cannot unambiguously make these distinctions, there is little hope that judges and juries can do better.

Although not all software patents suffer from abstract or overly broad claims, software technology is especially prone to these problems. Indeed, software patents are much more likely than other patents to have their claim construction reviewed on appeal—an implicit indication that parties to lawsuits have fundamental uncertainty over the boundaries of these patents. This uncertainty leads to more frequent litigation and substantially higher litigation costs.

Software patents are not just like other patents. The problems of software patents—problems arising partly from the nature of the technology

and partly from the way the courts have treated this technology—are a substantial factor in the overall poor performance of the patent system. The problem of implementing patent law to deal with abstract patents appears to be particularly stubborn and is unlikely to go away unless it is addressed directly.

Making Patents Work Like Property

What, then, will it take to fix the failure of patent notice and make the patent system an effective tool for encouraging innovation in all industries? At first glance, this might not seem too difficult a task, given that patents seem to have performed reasonably well as recently as the 1980s. Indeed, many people have been quite optimistic that the current round of draft legislation and recently renewed attention from the Supreme Court will soon lead to a rebirth of effective patent policy.

In chapter 10, however, we suggest that effective reform might well prove surprisingly difficult to achieve. Many reform advocates believe that the poor performance of the patent system flows from deterioration of patent "quality" (narrowly defined) that can be fixed by improving the patent examination process. We agree that invalid patents are a problem, and that patent examination can be improved; however, we see this as only *part* of the problem. We suspect that many people focus on patent quality because there has been so much publicity about bad patents on inventions that lack novelty or seem obvious, such as those mentioned previously involving the peanut butter and jelly sandwich or the backyard swing. Patents of doubtful validity create social costs, but our evidence suggests that concerns about validity are not the main drivers of the patent litigation "explosion."

Moreover, we think that attempts to improve patent quality, including review procedures involving third parties, will not be very effective unless there are broader improvements in patent notice. This is because patent examination *depends* on clear, predictable patent boundaries. For example, critics of the E-Data ruling contend that e-commerce had been discussed and practiced before Freeny's invention. Under patent law this "prior art"

should have invalidated the E-Data patent. But if the patent examiner, like almost everyone else, interpreted the patent narrowly in 1983 as claiming only in-store kiosks and vending machines, not e-commerce, then that prior art would not have seemed relevant. Thus, patent quality depends on well-defined patent notice, which involves much more than simply improving the examiners' access to prior art.

Finally, improving patent notice will be challenging because it cannot be achieved merely by a few court decisions or statutory changes; rather, it requires changing institutions. As we discussed above, the institutions of the patent system fail to perform basic functions required for notice—functions other property systems perform smoothly. Indeed, the institutions of the patent system actually seem to have contributed to the deterioration of notice over the last two decades.

In particular, the structure of the courts—specifically, the designation of a single court for patent appeals—appears to have undermined notice in at least two ways. First, a specialized court is more likely than a typical appellate court to take actions to expand its influence. This seems to have been the case with the changes in the interpretation of patent claims. The Federal Circuit downgraded the role of the Patent Office and the district courts in claim interpretation while increasing its own. We will show that this shift has decreased the predictability of patent boundaries. The Federal Circuit has also increased its influence by expanding the range of patentable subject matter to include software, business methods, early-stage inventions, and more.[18] Increased patenting of these new technologies might have created problems because of a second institutional shortcoming: a single appellate court might not be well suited for developing new law. Because power is concentrated in the Federal Circuit, patent law misses the benefits of the intercircuit competition that exists in most other areas of federal law.

We thus think it likely that effective reform will require structural changes, including, possibly, multiple appellate courts, specialized district courts, and greater deference to fact-finders. What other changes might improve patent notice? In chapter 11 we consider many reforms, most of which have also been advanced by others who have preceded us. These include:

- *Make patent claims transparent*. We recommend changes in the way patent claims are defined, published, recorded in the application

process, and used for subsequent determinations so that innovators have clear, accessible, and predictable information on patent boundaries. This includes strong limits on patent "continuations," a procedure used to keep patent claims hidden from the public for extended periods. We also consider a new role for the Patent Office where, for a fee, innovators can obtain opinion letters on whether their technology infringes a patent.

- *Make claims clear and unambiguous by enforcing strong limits against vague or overly abstract claims*. This includes a robust "indefiniteness" standard that invalidates patent claims that can be plausibly interpreted in multiple, fundamentally different ways. Also, we recommend reforms to limit overly abstract patents in software and other technologies. At the very least, patent law should prevent software patents from claiming technologies far beyond what was actually disclosed as the patented invention. If this proves inadequate, then we suggest subject-matter tests to limit the range of software inventions that can be patented, tests similar to those used during the 1970s and 1980s.
- *Make patent search feasible by reducing the flood of patents*. This includes a strong requirement that patents should not be granted on obvious inventions, coupled with substantially higher renewal fees. Ideally, patent renewal fees should be set by a quasi-independent agency and should be based on empirical economic research. These reforms will help stem the patent flood by screening out unwarranted patents and discouraging renewal of low-value patents. Reducing the number of such patents should help notice by reducing the cost of clearance search.
- Besides improving notice, we also favor reforms to mitigate the harm caused by poor notice. These include an exemption from penalties when the infringing technology was independently invented, as well as changes in patent remedies that might discourage opportunistic lawsuits.

In presenting these policy suggestions, we admit that we cannot know with any certainty what it will take to substantially improve patent notice.

These policy reforms move us in the direction of an effective patent system, but we cannot as yet tell whether they are sufficient to get us there.

Some have argued strongly that our policy prescriptions will not suffice. Economists Boldrin and Levine (2007) argue that the patent system does not work at all and should be abolished. We doubt that such an extreme move is warranted. Our evidence suggests that the patent system *does* provide positive innovation incentives for small inventors, and, on a larger scale, for chemical and pharmaceutical inventions. It seems likely that reform can improve notice and overall patent performance in some areas, especially since the patent system did provide positive innovation incentives as recently as the 1980s.[19]

On the other hand, we are troubled by the expansionist view of the courts that "everything under the sun made by man" should be patentable, including software, business methods, and even mental correlations. Tangible property systems are not so expansive. They restrict property to assets that can be clearly defined with unambiguous boundaries. A landowner receives no rights to untapped oil flowing beneath her land, or to migratory ducks that set down on it, or to the airplanes that fly over it. Similarly, we doubt that all types of inventive ideas can have clear boundaries, and our empirical evidence shows that many software and business-method patents fail to provide efficient notice. We are quite sure that the patent system needs to recognize the limits to its grasp, even if we are not sure of the best way to implement those limits.

Perhaps many of these reforms are not politically feasible today. Perhaps the political will to thoroughly fix patent notice does not yet exist. The patent bar has long dominated patent policymaking, and its interests—at least in the short run—do not always coincide with improved notice.

Yet there is some reason to think that this impasse is temporary and that some of the patent bar's opposition to improved notice will prove to be shortsighted. Our estimates suggest that the litigation burden imposed by patents is growing, and the performance of the patent system will continue to deteriorate. Moreover, the trends suggest that the deterioration might be particularly bad for software patents and other patents used in information technology (IT) industries—not only is the rate of litigation per software patent rising, but the share of software patents out of total patents continues to grow rapidly. It is no accident

that computer, semiconductor, software, Internet, and finance companies have begun to lobby for patent reform.

If this prognosis is correct, then the political landscape will continue to change. In the end, the survival of the patent system will require major improvements in the notice function. Despite all the rhetoric calling for "protection of inventors' property," today's patents fail as property, and tomorrow's might yet do even worse. Too often, such rhetoric is used to justify policies that actually undermine the property nature of patents. We hope our message and our empirical evidence succeeds in distinguishing actual patent performance from rhetoric. But in the long run, the pressures of market competition will determine the fate of the patent system based on its performance. If patents fail to work as property, over time they will make the United States economically less competitive, and industry will demand change.

Note to the reader: We wrote this book for a non-specialist audience, however, in several chapters we added sidebars on some advanced topics that might interest economists or lawyers. The main text can be understood without reading these.