

Innovation and Its Discontents

Adam B. Jaffe and Josh Lerner¹

Over the course of the nineteenth and twentieth centuries, the United States evolved from a colonial backwater to become the pre-eminent economic and technological power of the world. The foundation of this evolution was the systematic exploitation and application of technology to economic problems: initially agriculture, transportation, communication and the manufacture of goods, and then later health care, information technology, and virtually every aspect of modern life.

From the beginning of the republic, the patent system has played a key role in this evolution. It provided economic rewards as an incentive to invention, creating a somewhat protected economic environment in which innovators can nurture and develop their creations into commercially viable products. Based in the Constitution itself, and codified in roughly its modern form in 1836, the patent system was an essential aspect of the legal framework in which inventions from Edison's light bulb and the Wright brothers' airplane to the cell phone and Prozac were developed.

In the last two decades, however, the role of patents in the U.S. innovation system has changed from fuel for the engine to sand in the gears. Two apparently mundane changes in patent law and policy have subtly but inexorably transformed the patent

¹Brandeis University and Harvard University. Both authors are affiliated with the National Bureau of Economic Research. We thank Scott Stern for helpful comments. This essay draws heavily on Jaffe and Lerner (2004). All errors are our own.

system from a shield that innovators could use to protect themselves, to a grenade that firms lob indiscriminately at their competitors, thereby increasing the cost and risk of innovation rather than decreasing it.

Examples of dysfunctional patent behavior have become staples of the business and popular press. They range from the amusing and economically irrelevant, to not-so-funny cases that seriously threaten important technologies in important industries:

- Patents on inventions that are trivially obvious, such as the “Method for Swinging on a Swing,” “invented” by a five-year-old, and “User Operated Amusement Apparatus for Kicking the User’s Buttocks” (“invented” by a supposed grown-up);
- Patents in areas new to patenting, but covering purported discoveries familiar to practitioners and academics alike, such as Amazon.com’s attempt to prevent Barnesandnoble.com from allowing customers to buy books with a single mouse-click, and a bright MBA student’s patents on an option-pricing formula published in the academic finance literature two decades earlier;
- Patents that have become weapons for firms to harass competitors, such as the decade-long effort by Rambus, a semiconductor designer, to control computer memory technology by making sure that a long string of patents, all derived from a single 1990 patent application, incorporated important features of an industry-wide standard developed through a voluntary industry standard-setting association.

To major recent policy studies by the Federal Trade Commission (U.S. Federal Trade Commission, 2003) and the Board on Science, Technology and Economic Policy of the National Research Council (Merrill, Levin and Myers, 2004, cited hereinafter as “STEP Report”) have recommended significant changes to address these issues. In this paper, we provide an overview of the issues and discuss possible changes to address the widely perceived shortcomings of the current system.²

They Fixed It, and Now It’s Broke

The origin of these pathologies goes back to 1982, when the process for judicial appeal of patent cases in the federal courts was changed, so that such appeals are now all heard by a single, specialized appeals court, rather than the twelve regional courts of appeal, as had previously been the case. And in the early 1990s, Congress changed the structure of fees and financing of the U.S. Patent and Trademark Office (PTO) itself, trying to turn it into a kind of service agency whose costs of operation are covered by fees paid by its clients (the patent applicants).

It is now apparent that these seemingly mundane procedural changes, taken together, have resulted in the most profound changes in U.S. patent policy and practice since 1836. The new court of appeals has interpreted patent law to make it easier to get patents, easier to enforce patents against others, easier to get large financial awards from such enforcement, and harder for those accused of infringing patents to challenge the patents’ validity. At roughly the same time, the new orientation of the patent office has combined with the court’s legal interpretations to make it much easier to get patents.

² These general issues are also discussed in previous papers in the IPE series, including Gallini and Scotchmer (Volume 2, 2002) and Hall, et al (Volume 4, 2004).

However complex the origins and motivations of these two Congressional actions, it is clear that no one sat down and decided that what the U.S. economy needed was to transform patents into much more potent legal weapons, while simultaneously making them much easier to get.

An unforeseen outcome has been an alarming growth in legal wrangling over patents. More worrisome still, the *risk* of being sued, and demands by patent holders for royalty payments to avoid being sued, are seen increasingly as major costs of bringing new products and processes to market. Thus the patent system—intended to foster and protect innovation—is generating waste and uncertainty that hinder and threaten the innovative process.

The growth in the sheer magnitude of the patent phenomenon can be seen in Figures One and Two. The weakening of examination standards and the increase in patent applications has led to a dramatic increase in the number of patents granted in the U.S. Figure 1 shows that the number of patents granted in the U.S., which increased at less than 1% per year from 1930 until 1982 (the year the CAFC was created), roughly tripled between 1983 and 2004 (from 62 thousand per year to 187 thousand per year, an annual rate of increase of about 5.4%). Applications, too, have ballooned, to the point that there are now about 375,000 per year.

While some of this increase appears to reflect real growth in innovation, it is clear that a large part of the increase is a response to the increased laxity of the PTO, which

grants a significantly larger fraction of the applications it receives than do its counterparts in Europe and Japan. Figure Two is more worrisome still, showing a dramatic and inexorable increase since the early 1990s in the rate of litigation around patents. The number of patent cases filed has doubled in a decade and continues to rise. And the cost of defending a patent suit has risen as well; a patent infringement allegation from a competitor can now mean legal fees in the millions. For an under-capitalized startup, this prospect creates an overwhelming pressure to settle even frivolous complaints. Consumers therefore have less access to new products—from lifesaving drugs to productivity-enhancing software—than would be the case if innovative companies were not distracted from innovation by litigation and fear of litigation.

Much public attention has focused on the expansion of patenting into areas where it was previously unimportant or non-existent, such as biotechnology, software and business methods. Indeed, some of the worst abuses are in these areas. But concern about specific technologies potentially masks the deeper, fundamental problem. The incentives in the system now encourage frivolous applications, cursory review of those applications by the PTO, and indiscriminate filing of patent infringement suits as a generic competitive weapon. To get the system back on track, the system must be changed so that its incentives discourage frivolous applications, encourage rigorous patent examination, and discourage patent litigation where there is not a true invention to protect.

But it is time to recognize that the accidental combination of strengthening the legal value of patents while reducing the rigor of patent examination has damaged the system. It's time for a recalibration. Since the source of the mess is the combination of easier success in the courts and at the PTO, our proposed recalibration addresses both of these venues.

Goals and Objectives

Before getting to recommendations, it is important to be clear about what it is that any reform of the patent system ought to be trying to accomplish. While different analysts of the patent landscape have emphasized different aspects of the patent policy problems, there is general agreement on broad goals for reform of the system:

*Improve patent quality.*³ “Patent quality” is, to some extent, in the eye of the beholder. Certainly, as illustrated by examples discussed above, people are getting patents for inventions that are not new and/or are obvious. One way to solve this, of course, would be to make it much harder to get a patent on anything. If we did that, the few patents that did issue would be of very high quality, in the sense of being very deserved by the applicant. But the objective of patent quality has to be more than just making sure bad patents *don't* issue. It has to include also making sure that inventors *do* get patents when they have a truly novel, non-obvious invention, that such patents are processed relatively quickly and reliably, and that once granted they provide an adequate property right to protect subsequent investment in the invention.

³ See STEP Report, pp 87-94.

Reduce uncertainty. The primary objective of reform should be to reduce the uncertainty that now pervades many aspects of the patent system. (Ironically, the only aspect of the patent process that has become more certain is the application process itself, as the ultimate granting of some patent from each original application has become almost a sure thing!) The sand in the gears of the innovation machine is that companies and individuals must constantly fear that their research and product development may come to naught, because someone is going to assert an as-yet unknown or untested patent against them. Further, when such an assertion of patent infringement is made, the uncertainty about the ability to defend against that assertion often leads either to abandonment of the allegedly infringing technology, or to an agreement to pay possibly unnecessary royalties.

Keep costs under control. The PTO currently spends roughly \$1 billion/year for its operations. Patent applicants spend several times that amount, and patent litigants billions more. These resources might be well spent, if they achieved a reasonably smoothly functioning system. But the system is not working well, and it is reasonable to wonder whether we need to invest more of society's resources in the patent process. Ideally, we might argue, the PTO's finances should be decoupled from the amount that it raises in the form of fees: rather, it should spend whatever it takes to ensure high-quality applications. Realistically, however, it is unlikely that dramatically increased resources are going to be available, particularly for the operation of the PTO itself. So we need to look for solutions that go beyond throwing money at the problem.

Some Simple Truths

The next step towards reform is to understand some basic realities about the innovation process.

Mistakes will always be with us

Patent examination is never going to be perfect. Examiners are human. More important, there is an essentially irreducible aspect of judgment in determining if an invention is truly new. After all, even young Albert Einstein faced challenges while assessing applications as a “Patent Examiner-Third Class” in the Swiss Patent Office (Clark, 1973). Therefore, we cannot hope to have a system in which no “bad” patents ever issue. What is important is to have a system with fewer bad patents. And, since there will always be mistakes, it is important to have a system that functions reasonably well despite the issuance of some bad patents.

Better examination will require more resources. At current application rates, it would be very expensive to give all patent applications an examination sufficiently thorough to reduce significantly the problems with bad patents being issued. Now, the patent system is important, so it is possible that spending several billion additional dollars on the PTO would be worthwhile for society. But this kind of dramatic increase in PTO resources does not seem very realistic in the current fiscal environment. Fortunately, it is also not necessary to expend the resources necessary to provide very reliable examination for *all* patent applications.

Much more chaff than wheat

The first step to understanding why greatly increasing the resources for examination is not the best solution to the problem is to understand that most patents are, and always will be, worthless and unimportant. This is not a feature of the patent office; it is a feature of the innovation process. It is partly due to the human tendency for us each to think that our ideas are better than other people think they are. But it also reflects a deeper attribute of the process of technological development: the significance of a new idea usually cannot be known when it is first developed, because that significance depends on subsequent developments, both technological and economic. Many, many, “good” ideas are patented that never actually turn out to be worth anything. It is not that they shouldn’t have been patented to begin with. It’s just that for every invention with lasting technological or economic significance, there will always be dozens or hundreds of ideas that seemed *potentially* worthwhile, but which eventually proved to be valueless.

The fact that almost all patents are ultimately worthless has an important implication for the “patent quality” problem. If most patents are doomed to be consigned to the dustbin of technological history, it can’t make sense to spend a lot of resources to make sure that they all receive very high quality examination before issuing. The legions of inventors and patent attorneys may not like to think about this, but for the vast majority of patent applications, it will simply never matter—either to the inventor, her employer, or competitors—whether the patent is allowed to issue or not.

Day in and day out, most of what patent examiners do is like what the officials do in the last minute of a football game where one team is already winning by 30 points. They go through the motions of making rulings, because rulings have to be made, but they don't matter to the outcome of the game. The key difference is that in the patent game, much of the time no one *knows* whether this particular case is going to matter or not. It is as if nobody--the officials, the players, and the coaches--have any idea of the score of the game, or if the game even matters. But they all take it seriously because there is some chance that the particular "game" they are playing will turn out (months or years later) to be important. For the ones that do turn out to be important, it will matter a lot if patents are granted that should have been. But for the others, there will never be important technological or economic consequences. And these "others" are the vast majority of all applications in the system.

"Rational Ignorance"

If careful examination is expensive, and the vast majority of patents will never matter to anyone, then it would be inefficient to expend society's resources on careful examination of all patent applications. In the colorful phrase of Mark Lemley (2002), we can think of the poor quality of patent examination as representing "Rational Ignorance," by which he means that society is rationally choosing to remain ignorant about which patents really should be granted by the PTO. Lemley argues that it is, in fact, reasonably efficient to simply accept that PTO examination will be of poor quality, and that the cases that really matter will have to be sorted out in the courts. Court cases are expensive, but

because only the small fraction of patents that matter will ever get litigated, Lemley argues that the cost of litigation is, overall, efficient.

We agree with Lemley that it would be inefficient to provide thorough examination for all applications at the current rate of patent application. We disagree, however, that the current situation is acceptably efficient. First, while the out-of-pocket cost of litigation may be tolerable, the intangible cost of a system with pervasive low-quality patents is much higher than just the cost of paying lawyers to file and defend patent cases. The uncertainty that the current system creates for all parties regarding who can legally use what technologies is a cost that is very hard to quantify, but is surely significant. Talk to anyone involved in trying to commercialize new technologies, and you are likely to hear complaints about the headaches and uncertainty created by overlapping patent claims. Further, this uncertainty undermines everyone's incentives to invest in new technology. From the perspective of society as a whole, the loss of new products and processes that never make it to market, or that gain a toehold and are then abandoned after a threatened patent fight, is much larger than the visible costs of patent litigation. And, fortunately, there are changes that could be made in the system that would improve patent quality without requiring dramatic increases in the resources used in the examination process.

Inventors respond to how the Patent Office behaves

The key to more efficient patent examination is to go beyond thinking about what patent examiners do, to consider how the nature of the examination process affects the

behavior of inventors and firms. To put it crudely, if the patent office allows bad patents to issue, this encourages people with bad applications to show up. While the increase in the rate of patent applications over the last two decades is driven by many factors, one important factor is the simple fact that it has gotten so much easier to get a patent, so applications that never would have been submitted before now look like they are worth a try. Conversely, if the PTO pretty consistently rejected applications for bad patents, people would understand that bad applications are a waste of time and money. While some people would still try—either because they aren't smart enough to *know* they have a bad application, or because they are willing to take a roll of the dice—the number of applications would likely be considerably fewer than it has been in recent years.

Consider, just to illustrate the idea, the following thought experiment. Suppose that the PTO could dramatically reduce the issuance of patents on obvious or non-novel inventions by doubling the amount of time that the examiner spent on the average application. If the rate of application were unaffected by this change, it would require an approximate doubling of the PTO budget, as twice as many examiners would be needed to handle the flow of applications in a reasonable period of time. But it is unlikely that the rate of application would be unaffected by a dramatic change in examination standards. It is hard to know how much the flow of applications would be affected. But if the number of applications made each year were cut in half, then this doubling of examiner effort per patent could be brought about with *no* increase in the overall PTO budget.

This hypothetical is not intended to suggest that the problem is that easy; it is only meant to illustrate how the incentives faced by inventors and firms affect the efficiency of the system. As the quality of patent examination has deteriorated, the incentive for submitting marginal patent applications increased. A vicious cycle has emerged in which bad examination increases the application rate, which in turn overwhelms the examiners, reducing examination quality further and feeding on itself. If tools could be found to *improve* patent quality, this feedback would operate in the other direction, reducing the application rate and freeing up resources to further improve quality.

Potential litigants respond to how the courts behave

When the CAFC issues rulings that increase the chance of the patentee prevailing in an infringement suit, the consequences of this change are not limited to possible changes in the outcome of specific cases. Such a change in perceived success probabilities changes what disputes are, in fact, litigated. Conversations with attorneys involved in patent disputes make clear that the CAFC's strengthening of the offensive and defensive weapons of the patentee has significantly increased patentees' willingness to bring suit. Similarly, the change has significantly decreased the willingness of accused infringers to fight, even when they believe that the patents being used to threaten them are not valid. Constraining the growth in litigation, and the uncertainty created for all innovators by the risk of suit, will require a change in these incentives.

Get information to flow into the PTO

Another important aspect of incentives has to do with information: who has it, and what do they do with it? Much of the information needed to decide if a given patent application should issue—particularly information about what related technologies already exist—is in the hands of competitors of the applicant, rather than in the hands of the PTO. And there are strong incentives for firms to share this information. If a competitor of mine has filed a patent application, the last thing I want to see is for them to be issued a patent on an application that would have been rejected if the PTO had known about my technology. I would thus have a strong incentive to provide this information, if only the PTO would give me an opportunity for input, and if taking advantage of such an opportunity does not create strategic disadvantages for me down the road. So creating opportunities of this sort is another way that the system could exploit the incentives of private parties in order to increase efficiency.

But lest we get overly excited about the beauty of incentives, it is important to recognize that private parties' reactions to the incentives they face can also gum up the works. In particular, any opportunity that we create for outsiders to provide the PTO with information that is adverse to their competitors' patent applications will be exploited opportunistically. That is, even in the case of "good" applications, if I can easily throw some kind of speed bump in my competitor's path, I will probably be happy to do so. This means that any change in procedures that makes it easier for competitors to intervene will, to some extent, increase the cost, uncertainty and delay for valid patent applications.

Ultimately, attention to incentives can mitigate, but not eliminate, the tradeoffs that must be made among the cost of the system, its reliability in terms of screening out bad applications, and the speed and certainty with which good applications are processed into issued patents. We could have a system that made very few mistakes, and issued valid patents quickly, but it would be a very expensive system to run, because it would require a lot of time by very experienced examiners. We could have a system that put so many hurdles in the path of an application that bad patents almost never issued, but without a lot of resources such a system would inevitably slow down or deny many valid applications. Or we can have the existing system, in which we make it so easy to get a patent that a lot of stuff gets through that shouldn't.

What we can't do is to weed out the trash without killing any good stuff, and accomplish this greatly improved sorting without expending more resources. But perfection need not be the enemy of the good. If we pay attention to the incentives that different reforms create for desirable and undesirable behavior, we can recalibrate the system to get a better balance between rapid approval of good applications and reliable rejection of bad ones, and do it without dramatically increased resources.

Building Blocks of Reform

There are three key conceptual pieces for thinking about patent policy reform:

1. Investigate ways to create incentives and opportunities for parties that have information about the novelty and obviousness of inventions to bring that information to the PTO when it is considering a patent grant.

2. Consider the possibility for multiple levels of review of patent applications, with the time and effort expended escalating as an application proceeds to higher levels, so that money is not wasted on unimportant patents, but sufficient care is taken to avoid mistakes where the stakes are high.
3. Address the role of in ruling on claims of patent invalidity based on the existence of prior art, so that parties threatened by invalid patents have a reasonable opportunity to make their case.

The first two of these concepts are aimed at making the PTO more effective at reasonable cost. The third addresses the reality that the best of all possible PTOs will still make mistakes, and so we need a court system that is capable of rectifying those mistakes. The next section considers the possible implementation of the changes at the PTO.

The Quest for Quality at the PTO

At a conceptual level, ensuring the quality of the PTO examination process has two key building blocks. First, Congress and the Courts must provide the PTO with an appropriate definition of the standard for issuing a patent, particularly a definition of non-obviousness that separates the wheat from the chaff. Second, the PTO must have the appropriate procedures to implement that standard.

A possible revamping of PTO procedures rests on the concept that the most efficient balancing of the need to bring in outside information against the reality that most patents are unimportant is brought about by a process with multiple potential levels of review. Examination could begin as it does now, with the review of an application by an examiner, and no participation by other parties. If, however, the examiner makes a determination that a patent should be allowed to issue, there could then be an opportunity for “*pre-grant opposition*.” A public notice of the intention to issue a patent would be followed by a brief period of time in which other parties could submit to the examiner evidence, if it exists, of “prior art” that they believe should be the basis for a finding that the invention is not novel or is obvious and hence should not issue. This pre-grant opposition would not give outside parties any opportunity to argue their case, and they would not have access to legal discovery processes to produce additional evidence of prior art. It would simply be an opportunity for parties that have information in their possession to put this information before the examiner.

If this option had been in place when certain problematic patents were being pursued—such as the award to MBA Vergil Daughtery for the “expirationless option”—it would have been a simple matter for an investment bank, or an annoyed academic, to send the patent examiner the 1960s-vintage papers on the subject by Paul Samuelson. While the articles themselves are somewhat technical, one only needs to read their abstracts to realize that an infinite-lived option is not a new idea. With these papers in his possession, the examiner would have known that the Daughtery application was not novel, and presumably would have rejected it.

If, after reviewing any evidence that arises in this manner, the examiner decides to issue the patent, there could then be a final opportunity for review in the form of a request for re-examination. This request would have to include a stated basis for a case that the patent is invalid, and the PTO could decline to grant the re-examination request if no such basis exists. But if re-examination does commence, it would be a complete review of the initial decision, undertaken by an independent examiner, and with opportunity for the party requesting re-examination to argue their case.⁴

The logic of re-examination—as distinct from pre-grant opposition—can be seen in the case of the Amazon “one click” patent. Unlike the infinite-lived option, there is no pre-existing published scientific paper that unambiguously demonstrates the existence of prior art rendering the one-click patent non-novel or obvious. Rather, the case for its obviousness lies in the broad pattern of software practices in use over time. Making the case that this pattern rendered the Amazon application obvious would probably require argument and explanation, perhaps including testimony by expert witnesses. By design, the pre-grant opposition procedure does not permit this, because it must be kept quick and simple so as not to delay the bulk of valid patent grants. But a tremendous amount was clearly at stake in this dispute. If a viable re-examination option had existed, barnesandnoble.com and other parties would have had the incentive and opportunity to

⁴ Both the FTC and the STEP Report advocate the initiation of some kind of post-grant re-examination process, though not necessarily with the specific features discussed here. For empirical analysis of the operation of post-grant review in Europe see Hall, et al (2004).

demonstrate the invalidity of the Amazon patent, resolving the dispute without the need to involve the courts.

The logic of this escalating series of examinations is that most patents would never receive anything other than the most basic examinations. But for those applications that really matter, parties would have an incentive and opportunities to bring information in their possession before the PTO, and the PTO would have the opportunity to make sure it makes the right decision in the cases that really matter. Let's now consider each of these steps in some detail.

Pre-grant opposition

The logic of allowing limited opposition prior to grant is that the PTO cannot know everything there is to know, and the applicant does not have appropriate incentives to bring in information that undermines validity. Other parties *are* likely to have information that bears on validity, and they have an incentive to provide it. They have not historically been given any opportunity to provide it, partially because patent applications have been secret up until the time that a patent is granted. But now that most applications are published 18 months after application, there is really no reason not to allow parties that have information that they believe bears on the validity of a patent application to bring that information to the examiner's attention. At this stage, only the examiner would decide whether the information is relevant, and what consequences, if any, it has for the application. Therefore, there need not be any significant delay or increased expense.

Under current procedures, outside parties do have certain opportunities to insert information about prior art into the PTO process, but these opportunities are structured in such a way as to minimize both the incentive to use them, and the possibility that they will actually affect the outcome. For example, in the case of continuation and divisional applications, it is possible to learn from the PTO website that such an application has been filed, even before the application is legally “published.” Under PTO rules, one can file a “protest” of such a pending application, arguing that the application should not be granted, on any grounds related to patentability, including obviousness. But the catch—which has no apparent purpose other than making it hard to get outside information into the process—is that such a protest has to be filed *before* the application is legally published.⁵ Since legal publication occurs eighteen months after filing, this puts the would-be protester in the odd position of watching for new applications to pop up in the list of divisional/continuation applications, and getting in the protest before the application is technically “published.” Go figure.

The other existing mechanism for adding information to the PTO record (short of requesting re-examination, discussed further below) is to simply dump citations of pre-existing patents or other publications into the patent file *after the patent has issued*.⁶ Of course, this is too late for the information to have the desired beneficial effect of getting the examiner to make the right decision before it is too late.

⁵35 USC 122(c); 37 CFR § 1.291.

⁶35 U.S.C. § 301.

The most efficacious time period to allow outsiders to provide information about prior art is, of course, *after* publication—when the world is officially on notice that the PTO is considering granting a patent—but *before* the patent is granted, so that the information can be considered by the examiner before making that decision. As we have emphasized, creating such an opportunity need not unduly gum up the process, because outside parties would not be afforded the opportunity to argue their case or otherwise participate in the proceeding—only to provide information.

Given this limited participation, it is important to consider how the legal treatment of information so provided affects the incentive to provide it. In particular, if the patent issues despite the purported prior art submitted by an outside party, a legal presumption regarding the validity of the patent over the art that was provided by the outside party would seriously undermine the incentive to provide information in this way: someone considering giving the examiner information would have to worry about “wasting” good stuff on an examiner who might not understand it, thereby destroying or seriously weakening its value in some future forum.

To make this concrete, suppose this procedure had been in place while the Dougherty patent was being examined, and someone had sent in the Samuelson articles, but for whatever reason the examiner still issued the patent. If there is ever litigation over this patent, parties would want to challenge the validity of the patent on the grounds that Dougherty’s invention was not novel, given the previously published paper. If these

papers had indeed been put in front of the examiner, standard legal practice would be to presume that they did not invalidate the patent, since the examiner saw them and issued the patent anyway. This presumption would make it more difficult to invalidate the patent on these grounds. If the party supplying these papers had had an appropriate opportunity to explain their significance to the examiner, it would be reasonable to presume that an examiner that nonetheless ignored them had good basis for doing so, and to put a heavy burden on anyone who later wants to argue their relevance. But the proposed procedure does *not* give the provider of such information any opportunity to explain its significance, so there is no reason to create a presumption that the information has been *appropriately* considered. Hence it is perhaps appropriate, and important in terms of maintaining good incentives, to allow people to submit such information, but also to use it later, if necessary, without any adverse presumption.

Post-grant re-examination

The logic of making any opposition that occurs prior to the patent grant be of the limited form described above is so that it cannot be used strategically by competitors to delay or obstruct the granting of a valid patent. But because of this limited nature, pre-grant opposition is not likely to be sufficient to ensure that invalid patents are not issued. For this reason, it is worth considering an additional step in the escalating intensity of examination, in which there is an opportunity for formal re-examination of issued patents. Compared to pre-grant opposition, re-examination should afford parties challenging a patent more opportunity to make their case. But because of the possibility of

opportunistic opposition to valid patents, it is worth considering putting barriers in place that limit firms' ability to use re-examination strategically or frivolously.

The use of re-examination to eliminate bad patents exploits two of the important Simple Truths. First, it focuses additional examination resources not on *all* patents, but on the relatively small fraction of patents that are important enough to care about. It thereby offers hope to improve examination standards in a cost-effective manner. Second, it creates an opportunity for outside parties that hold relevant information about patentability to bring that information to the PTO's attention. By creating an incentive for outsiders to bring the relevant information to the examiner's attention, the examiner can make a better decision while using fewer public resources than would be necessary if the examiner had to go out and find all relevant information herself.

A final benefit of workable pre-grant opposition and re-examination procedures is that they would assist the PTO in getting itself educated about new and rapidly evolving technologies. Giving outsiders the opportunity and incentive to bring important, relevant information into the process will not only improve decisions in the individual cases where it occurs, it will provide general education for examiners about how new, dynamic fields are evolving, and where the prior art tends to be found. This should result in better decisions across the board.

Some strengthening of the examination received by all patents could complement the institution of effective re-examination and pre-grant opposition. The Rational

Ignorance principle means that it is not efficient to give all patents the kind of scrutiny that they get in litigation or even in a re-examination. But the current standard is so low that it is almost surely inefficient. If the PTO set as a goal to approximately double the average examiner time per granted patent—including whatever time is needed to review evidence submitted under pre-grant opposition—and *combined* this increase care with effective re-examination, it ought to be able to credibly commit to a significant reduction in the issuance of obvious and non-novel patents. This could be expected to set up the “virtuous cycle” by discouraging applications that are made today only because applicants know that they can get away with it.

Unfortunately, there is no way to know how large a reduction in applications would be brought about by a credible tightening of examination standards. But there is reason to believe it would be significant. Since 1990, when the PTO was converted to its current user-friendly structure, applications have increased from about 150,000 per year to about 350,000 per year. There are multiple factors at work in this increase, including the enhanced value of patent protection since the creation of the CAFC and the expansion of patentability to cover unambiguously the areas of biotechnology, software and business methods. There is also some evidence that the increased rate of application represents a real, broad increase in the underlying innovation rate (Kortum and Lerner, 1998 and 2003). But it is unlikely that these factors explain a greater than 100% increase over a dozen years. It is likely that a transformation of the examination process such that applicants know that frivolous applications will be denied would reduce the application rate to the range of 250,000 per year, roughly a third less than the current rate, and the

rate that prevailed as recently as 1998. This means that a doubling of examiner effort per application could be brought about with only about a one-third increase in examiner resources. (Devoting twice as much effort to 250,000 applications would be equivalent to maintaining the current effort level if there were 500,000 applications, which is about a third more than there are now.)

Whatever the numbers might turn out to be, the important principle is that the loosening of examination standards in the last decade has set up a vicious cycle that is not good for inventors or for the PTO. This cycle has to be reversed. It will take time, because potential applicants will not immediately change their behavior. But the PTO must undertake real reform in order to break the vicious cycle, and try to establish the virtuous cycle of more credibly discouraging frivolous applications, which in turn will make it easier to muster the resources to sustain credibility.

Devilish details

Patent opposition and re-examination are not, in fact, new ideas. European countries have used patent oppositions for many years (Graham, *et al.*, 2002; Hall, *et al.*, 2004). Congress first introduced re-examination into the U.S. patent system in 1980. This procedure did not, however, allow parties other than the PTO and the applicant to participate in the proceeding, thus missing the opportunity to take advantage of information in the hands of third parties. In 1992, the Advisory Commission on Patent Law Reform (1992) recommended that the re-examination procedure be expanded to

allow for the participation of outside parties. In 1999, Congress introduced such 3rd-party participation as part of the American Inventors Protection Act (“AIPA”).

Re-examination under the AIPA has not been a success. Only 300 patents were, in fact, re-examined in 2002. And of these, only *twenty-five* were re-examined under the new AIPA procedure that permits third parties to participate in the re-examination. Given that some amount of mistakes in initial examination is inevitable, even a PTO operating efficiently and appropriately would likely have generated more than 300 re-examinations from the approximately 190,000 patents currently granted per year. Given the general agreement that a lot of these grants are dubious, it is inconceivable that a re-examination procedure perceived to be fair and effective would have gotten so little business.

The reasons why the new procedure is not widely used are easy to understand: Congress ignored some key recommendations of the 1992 Commission when it created the rules for re-examination.

The tension in the design of this process derives from the incentives—good and bad—that it creates. We want a procedure that makes it feasible and attractive for outside parties to bring to bear real information that they possess about the validity of a given patent. But we *don't* want to create an opportunity for competitors to frivolously or maliciously try to shoot down valid patents. Given that validity and frivolity are always to some extent subjective, these two objectives are partially in conflict.

In passing the AIPA, Congress erred on the side of making sure that the re-examination could not be abused to hold up valid patents. But the protections that it built in for the patentee make the procedure very unattractive, even to a party that has a valid basis for challenging a patent:

- First, if the re-examination results in the patent grant being withdrawn, the applicant can then appeal this decision to the courts. But a decision *not* to withdraw the grant cannot be appealed by a challenging party. This makes the risks inherent in the process asymmetric.
- On top of this, if there is ever subsequent litigation over the patent—because, for example, the patentee sues the party that challenged the patent for patent infringement—the challenger is legally barred from making any argument regarding the validity of the patent that they *could have* made in the re-examination, even if that argument was never considered by the PTO.
- Finally, the kind of evidence that can be brought by third parties to try to prove invalidity has been limited in two important ways. First, the evidence must be in the form of patents and other printed publications. For many of the most controversial patents, particularly in the software and business methods area, the evidence regarding the existence of “prior art” that ought to invalidate the patent may not be in the form of patents or other printed publications. Second, until this portion of the law was changed in 2003, challengers were barred from presenting at the re-examination evidence that was considered by the examiner during the initial examination process. This means that the re-examination process was

useless for the situation where an examiner saw a piece of prior art, but failed to grasp its significance.⁷

So suppose that your competitor has just gotten a patent that you think is bogus, but which you are worried could be claimed to cover one of your products. You can ask for re-examination, but if you do, your hands will be tied as to what evidence you can bring; if you win the competitor can appeal, but if you lose you can't; and if you lose and end up in litigation later, you will be barred from making any argument in the course of that lawsuit that the judge decides you *could have* brought in the re-examination. Nine times out of ten, asking for re-examination is not going to be an attractive bet. You are better off waiting and taking your chances in court.

To create more balanced incentives for outside parties to come forward with information that they have regarding validity, the re-examination process could have the following features:⁸

1. Parties could be able to bring forward any relevant factual evidence.

⁷The prohibition on evidence previously considered by the examiner derived from a CAFC decision under the pre-AIPA procedure in which re-examination could be initiated by the PTO but excluded third parties (*In re Portola Packaging*, 110 F 3d 786 (Fed. Cir. 1997)). Congress could have overruled this decision in the AIPA, but chose not to. A law passed by Congress and signed by President Bush in 2003 finally overturned this decision.

⁸Table 4-1 in the STEP Report provides a useful summary of important features of a re-examination procedure.

2. If a patent survives re-examination, parties could be barred in subsequent litigation only from making arguments that were specifically made and rejected by the PTO in the re-examination.
3. Re-examinations could be conducted by a specialized group of “re-examiners.” An examiner who is drawn from the same group as the original examiner cannot be expected to have an open mind about whether a mistake might have been made.
4. Both the patentee and the challenger could be given the right to appeal the PTO’s decision.

These changes would shift the balance between encouraging valid challenges and discouraging frivolous ones in favor of more challenges. These changes can be prevented from stimulating *excessive* challenges by appropriate countervailing incentives. The AIPA already requires that the PTO find that a substantial new question of patentability has been presented, or it does not initiate a proceeding. This provides for quick and inexpensive disposition of truly frivolous challenges. Finally, the incentives of both applicants and potential challengers to avoid wasteful proceedings would be improved by the appropriate use of fees and cost-shifting. There could be a non-trivial fee for initiating a re-examination proceeding, say \$50,000, paid in the first instance by the party challenging the patent. If the challenge is successful, and the patent is revoked, then the original patent applicant could be required to pay this amount plus the challenger’s legal fees, to the challenger. Conversely, if the challenge is unsuccessful, the challenger could

be required to reimburse the applicant's costs for defending the patent (as well as having paid the fee for initiating the proceeding).⁹

For a potential challenger who truly believes that an invalid patent poses an important threat to her business, this fee, and the risk of paying both sides' costs, will not be a significant disincentive to bringing the challenge. But it will provide some deterrent to someone who hopes only to throw sand in the works—knowing that the patent will likely still be held valid in the end. Conversely, for applicants that know they have pulled the wool over the examiner's eyes, the prospect of paying the fee plus both sides' costs may seem like a high price to pay for merely delaying the likely withdrawal of the patent. They will therefore have an incentive to fold their tents and go away. Further, the knowledge that this expensive and unattractive prospect likely lies at the end of the road will discourage marginal applicants from filing patent applications in the first place.¹⁰

The other major aspect of reform that has been discussed by all observers is increased rigor of initial examination of patents. In June of 2002 the PTO released with some fanfare a "21st Century Strategic Plan" intended to transform itself into a "highly productive, quality-focused organization." It contains a variety of proposals designed to improve the functioning of the examination process. It also proposed an increase in

⁹ The STEP report advocates a fee for requesting re-examination, but does not contemplate fee-shifting based on the outcome (STEP Report, pp 97-103).

¹⁰ We should also note that as part of the AIPA, the U.S. Congress stipulated a two-month window in which others can submit prior art after patent applications are published. No one seems to do so, because (a) not all applications are published, (b) many parties are still unaware of this provision, and (c) people don't want to limit their ability to use prior art in subsequent litigation.

application fees, and called upon Congress to appropriate all of the fees collected for patent applications to the PTO, instead of funding the PTO at levels below the revenues collected, as has been the practice in recent years.

After discussions with the patent community, the plan has been changed somewhat, and aspects of it have been presented to Congress. One aspect is a PTO proposal to “outsource” the search of prior art to private companies, leaving examiners only with the final determinations regarding patentability, to be based on the prior art identified by private contractors. In the June 2002 version of the Plan, the PTO had proposed that the *applicants* be required to hire search firms to prepare a prior art review for the examiner. This fox-hiring-the-guard-for-the-chicken-coop approach was roundly criticized, leading the PTO to amend its proposal. Even with the PTO responsible for supervising the outside search firm, this proposal remains controversial. The organization that represents examiners opposes it, on the grounds that examiners cannot ultimately be responsible for the quality of the validity decision if they are not doing the prior art searching (Stern, 2003).

Ultimately, the decision as to whether the PTO search function remains “in-house” or is moved to the “outhouse” (so to speak) is not going to determine whether patent quality is improved. Until the process is changed so that other parties that know something about the technology surrounding a given application have the opportunity and incentive to bring that knowledge forward, there will be no cost-effective way to fix the problem of low quality patents.

Many players in the patent community have endorsed the PTO proposal to allow it to, in effect, retain all of the fees that it collects and spend those resources on the examination process (Intellectual, 2003; Kirk, 2003). It is likely, however, that higher fees, coupled with a truly significant improvement in examination standards, would reduce the number of applications. As discussed above, such a reduction is socially beneficial, because it would allow the PTO to concentrate its resources more effectively on the applications it does get. But it also means that, if the plan is truly successful, the total revenue collected will not be as great as the PTO is anticipating. Indeed, revenues could easily fall *below* the level of expenditures. When that happens, the PTO may regret having insisted on the importance of the link between PTO revenues and PTO expenditures.

This discussion illustrates an important point that is not visible in the current debate about the PTO Strategic Plan. The debate over “revenue diversion” is fundamentally off the point. It presumes that the level of fees that the PTO collects ought to determine the amount of money that it gets to spend. But that is backwards. Economic analysis would say that Congress should start by figuring out how much money the PTO needs to do its job right. Because of the need to train and retain more and better examiners, and to update its information systems, that amount is probably somewhat greater than the amount that the PTO is being allocated. This need—not the fact that appropriations are less than revenues—is the economically supportable reason why PTO appropriations should increase. Now, it is not unreasonable for Congress to

take the position that the cost of running the PTO ought to ultimately be borne by patentees. This implies that the increased appropriations should be matched by PTO revenues, at least eventually, but by this analysis the link runs from needed appropriations toward the setting of revenue levels—not the other way round.

Of course, giving the PTO more money will not magically make it more effective. It must also solve the management failures that plagued all of its earlier efforts to modernize its operations. More fundamentally, it has to change the very concept of “productivity” that it pursues. While the 21st Century Plan makes the right noises about “quality” rather than just serving “customers,” the fact remains that the PTO defines its management objectives in terms of reducing the time it takes to process patents, and continues to reward examiners based on measures of productivity that encourage granting patents rather than granting only valid patents. What you measure and what you reward is going to be what your employees deliver. If the PTO is serious about patent quality, it has to overhaul its compensation structure, so that examiners are rewarded for denying applications on non-novel inventions, and for making those denials stick. Otherwise, the vicious cycle will continue, no matter how much more money is spent.

Leveling the Judicial Playing Field

Breaking the vicious cycle of bad examination and bad patent applications is the key to reform of the process that produces patents. But as we have emphasized, there are always going to be mistakes, and so it is important that the court system operate as efficiently as possible to rectify those mistakes, while also permitting owners of valid

patents to enforce the legal rights the patent conveys. As discussed above, the CAFC has significantly tilted the legal playing field in favor of patentees. In this section we discuss possibilities for adjusting this tilt, to preserve the rights of holders of valid patents while improving the reliability with which bad patents are weeded out.

The presumption of validity

The problem of actual or threatened infringement suits based on dubious patents is greatly aggravated by the legal doctrine that a patent granted by the patent office is entitled to a legal presumption as to its validity. Because of this presumption of validity, anyone challenging an issued patent must prove by “clear and convincing evidence” that the patent is invalid. The “clear and convincing evidence” standard is not as high as the “beyond a reasonable doubt” that must be proved in a criminal case, but it does tilt the legal playing field in a validity dispute in favor of patentees. (By way of contrast, to win its infringement claims the patentee must prove infringement only by a “preponderance of the evidence,” meaning that the balance of the evidence is in its favor.) Given that issues of prior art and obviousness are inherently to some degree subjective, setting this high standard for proof makes it hard for anyone to be confident that they can invalidate a patent, even if they think they have pretty good evidence of, for example, having independently developed their technology before the patented invention.

Another aspect of the presumption of validity is that the kind of evidence that can be presented to prove invalidity is limited. If I am trying to convince a jury that a patent was granted that should not have been, I might want to try to show just how few hours

the examiner worked on the application before granting it, or the number of other patents approved by this examiner that have subsequently been found to be invalid. But such evidence is typically not allowed, on the theory that an examiner working in her official governmental capacity has to be presumed to have done their job appropriately.

These legal rules go a long way to explaining why many firms, faced with a claim of infringement of a patent that they think is invalid, nonetheless will drop an infringing product or pay a royalty. It simply may not make sense to fight if some of your weapons are inoperable, and your opponent is protected by high walls.

The presumption of validity accorded patents has an explicit statutory basis,¹¹ but it is an example of a broader principle of administrative law: issues that have been appropriately vetted before a competent decision-making body should be presumed to have been decided correctly. From the local zoning board to the U.S. Environmental Protection Agency to the Federal Trade Commission, these decisions are typically made after some kind of open administrative process, in which all interested parties generally have a right to participate. There is an inherent logic to affording a degree of deference to decisions made in this manner.

But as we have seen, the process by which a patent is granted is fundamentally different from these other administrative decisions. All interested parties are most definitely *not* invited to participate in the examination process. There is thus

¹¹ 35 U.S.C. § 282

fundamentally a much weaker logical case for the presumption of validity that the patent statute affords to issued patents.

Going beyond the conceptual framework, the current practice of the PTO is also clearly inconsistent with a presumption of validity. Outside of legal doctrine, reasonable people do not hold a presumption if everyday observation demonstrates that the presumption in question is often false. Further, because of the Rational Ignorance principle, there is a sense in which validity could never be a logically sensible presumption to make about all patents. So it might seem logical to drop the presumption of validity, allow evidence of examiner indifference or incompetence to be presented if it exists, and create a level playing field on which the jury simply decides whether the evidence, on balance, favors validity or invalidity of a challenged patent. Indeed, the FTC in its report specifically recommends Congressional action to change the standard of proof of invalidity to the balanced “preponderance of the evidence” standard (U.S. F.T.C., 2003, p 8).

There is, however, an important reason to maintain the presumption of validity. Remember that the fundamental purpose of the patent system is to give inventors a basis for expecting that they will have an opportunity to recover investments that they make in developing and commercializing their invention. When a start-up firm goes out to raise money for this purpose, it is important that the patent or patents that are claimed as the basis for the protecting the firm’s technology have the presumption of validity. If, instead, the validity issue were reduced to a legal coin flip, it would greatly increase

uncertainty. Uncertainty is the enemy of investment, so patents of uncertain validity would be much less effective in providing a base for development of innovations.

For this reason, eliminating the presumption of validity is a potentially dangerous change in terms of its long-run consequences for the innovation process. The alternative is to change the system so that it is, in fact, a reasonable presumption to hold. This is why a viable re-examination process is so important. Because of the Rational Ignorance principle, it would never be reasonable to assume that the output of the overall initial examinations process could be presumed to be valid. But if all parties have the opportunity to request re-examination on the basis of factual evidence in their possession, then the presumption might become reasonable.

If re-examination was never requested on a given patent, it is indeed reasonable to presume that the patent is valid, because the parties most likely to hold evidence of its invalidity had an incentive and an opportunity to present that evidence. This is not to say that a patent for which re-examination was never requested is *proven* valid, but only that it is reasonable for there to be a presumption of validity, with all that implies regarding the standard of proof that must be met by a party that ultimately does wish to challenge such a patent. And if a patent is re-examined, and survives re-examination, the Rational Ignorance principle does not apply. A request for re-examination—particularly if it is a reasonably expensive request to make—combined with a decision by the applicant not to withdraw in the face of such a request, tells the PTO that this is an important application. We should expect the patent office in a re-examination proceeding to devote sufficient

resources to “get it right.” Again, this procedure doesn’t prove validity, but it is enough to form a basis for a presumption of validity.

Thus the existence of a viable re-examination option serves the interests both of parties worried about invalid patents and parties who want the full economic benefit of their valid patents. It helps the former by providing a forum in which appropriate incentives are created for third parties to bring forth relevant facts, and for the PTO to devote the appropriate resources to sifting through those facts. Perhaps less obviously, it helps the holders of truly valid patents, because it can be the “dog that didn’t bark.” If an effective re-examination procedure exists, then the fact that it is *not* invoked in a given case provides a logical (as distinct from a legal) basis for overcoming the Rational Ignorance principle, and truly presuming that the patent is valid.

Trial by jury

Another complaint of attorneys who defend infringement suits is that the right of jury trial stacks the deck in favor of patent holders. There seem to be several parts to this argument. First is the uncontroversial observation that the evidence in a patent case can be highly technical, and the average juror has little competence to understand and evaluate it. On the surface, the effect of juror incompetence would seem to be neutral as between an accused infringer and a patentee. Even if it is neutral, however, having decisions made by people who can’t really understand the evidence does increase the uncertainty surrounding the outcome. Such uncertainty is a major factor in accused infringers settling rather than fighting even when they think they have a pretty good case.

More subtly, jurors' inability to grasp technical evidence may interact with the presumption of validity in a way that helps patentees and hurts accused infringers. Where the standard of proof is that whichever party on balance presented better evidence wins, then perplexed jurors would not seem to favor one side over the other. And this is, indeed, the standard for proving infringement, which is what the patentee needs to prove to win. But when one side has to achieve a reasonably high level of proof, it seems plausible that jurors' inability to truly understand the evidence being presented is not neutral, but rather acts against the party that must achieve a high standard of proof. If, at the end of the trial, the jurors are simply befuddled by the evidence, the most likely outcome is that they will conclude that neither side has made a convincing case. Thus the "clear and convincing evidence" standard *combined with* decision-making by juries makes it likely that the patentee will win on validity questions. (Which is why most patentees ask for a jury trial!) Put another way, it is very difficult to ever make the evidence "clear and convincing" to a group of people who do not have the necessary training and education to understand it.

Regardless of whom it benefits, to non-lawyers it does seem hard to argue that lay jurors are the best decision-makers in patent suits. Of course, the right to a jury of one's peers is a venerated concept in Anglo-American law. But there isn't really any sense in which a patent jury is, in fact, a jury of peers. A jury of scientists and engineers—the actual peers of the inventor—probably would be a relatively competent decision-making group. But, of course, that is not what we get. If we left patent cases to judges rather

than juries, we would still not have scientists. But judges spend their professional lives evaluating evidence in many different disciplines, and have to develop some ability to sort through it. Further, a judge always has the ability to appoint a “Master,” an outside expert in the service of the court, who can rule on specific technical questions put to her by the judge.

The CAFC has, in fact, put some limits on the role of juries in patent cases. In particular, it is the job of the judge, not the jury, to interpret the patent’s claims.¹² Typically, judges receive written and oral arguments, and often the testimony of expert witnesses, and then issue detailed instructions to the jury regarding what the claims mean. Thus, when it comes to the question of infringement—on which the patentee needs to carry the burden of proof in order to win—the judge assists the jury by interpreting the technical language of the claims before putting the question of infringement to the jury. But when it comes to a lack of novelty or obviousness—which the accused infringer must prove in order to invalidate the patent, and must do so with clear and convincing evidence—the jury gets no such help. These decisions are inherently no less technical than those of claim construction, and there does not appear to be any logical or substantive argument as to why a lay jury is the appropriate decision-making body for these questions. It would be entirely feasible for the judge to “construe” the novelty and obviousness of the patented invention relative to some other invention, just as the judge “construes” the claims of the patents. The ultimate question of patent validity could still be left up to the jury. If, for example, the judge ruled that the patent at issue was obvious

¹²*Markman v. Westview Instruments, Inc.*, 52 F. 3d 967 (Fed Cir 1995).

or non-novel relative to some other specified invention, there might still be a dispute as to whether that other invention was, in fact, part of the prior art at the time of patent filing. The jury could then appropriately decide that question, based on testimony and documentary evidence as to when the invention in question came to be known.

If the PTO were revamped so that a presumption of validity was appropriate, and if the burden of proving invalidity by “clear and convincing evidence” were made feasible by removing the technical determination of novelty and non-obviousness from the jury, then patent litigation would be the appropriate last resort when disputes over patent claims cannot be resolved any other way. There would still be patent suits, and they would still be expensive. In areas where technology is changing rapidly, and there are numerous competing and overlapping claims, there would still be considerable uncertainty about who has the rights to what technology. But the pervasive fear that almost any modern (or not so modern!) product or process is continuously at risk of facing an infringement claim would be dramatically reduced. And when claims are made based on patents of questionable validity, accused infringers would negotiate from a position where both parties expect a reasonably competent determination as to novelty and non-obviousness. This should reduce (though not eliminate) the incentive to pay royalties and settle rather than undertake a challenge that is risky, no matter how questionable the validity of the asserted patent.¹³

¹³ The STEP report recommends several specific changes to modify or remove certain “subjective” elements of patent litigation, which the STEP Committee concluded increase the cost and uncertainty of patent litigation. See STEP Report, pp 117-123.

Software, Genes, and Other Alleged Patent Nightmares

We have seen the difficulties that have been created by bad patents in many different technologies and industries. And we have discussed the inherent difficulties that arise because granting patents restricts and inhibits cumulative and overlapping inventions. The possible changes described above are intended to create a patent system that better encourages invention, though we have no illusion that any reform can create a perfect system, or one that somehow eliminates the inherent tradeoffs.

Other observers, grappling with the current dysfunction of the patent system with regard to particular technologies or industries, have concluded that the problems of encouraging invention in that particular setting require a patent policy that distinguishes among technologies. Software, business methods, and certain aspects of biotechnology such as genetic sequences are all technologies for which the courts has expanded the range of patentable subject matter beyond what was perceived to be patentable at the end of the 1970s. Each of these areas has subsequently been characterized by major controversies over patents that appear to be either invalid, overly broad, or both, leading to concern that the patent system is inhibiting rather than encouraging invention in these areas. As a result, there have been numerous suggestions that inventions in these areas should not be patentable, or, if patent protection is to remain available, that different rules and procedures are necessary to adapt the institution of patents to these technologies.

Theoretical arguments can certainly be made in favor of “tuning” the attributes of patent protection in different technological areas to reflect the attributes of different

technologies, including eliminating patent protection entirely for technologies with certain attributes. But opening the door to such tuning is likely to lead quickly to special pleading that will not serve the public interest. To understand these arguments, we consider in turn the three most important perceived problem areas: business methods, software, and biotechnology.

Funny business over business methods

The CAFC has eliminated the long-perceived prohibition on patenting business methods, and there have been some silly manifestations of this new art form, such as the Amazon one-click patent. The European Patent Office does not permit patents on business methods, and some have suggested that creating a new business method is not, fundamentally, an act of “invention,” and hence should not be the basis for a patent. Following the controversy over the “one-click” patent, Jeff Bezos of Amazon.com proposed that the patent life for software and business method patents be reduced from the standard 20 years to only 3 to 5 years.

It is clear that the PTO has done a disastrously bad job in testing applications for business patents against the prior art. And this is not really surprising. The prior art that patent examiners find most easily is previously issued patents. Beyond old patents, they have some ability to search for and identify prior art that appears in other published forms, such as scholarly publications. But in the area of business methods, most of the prior art is not in patents; indeed, it cannot be, because until recently people thought that business methods were unpatentable! And the prior art is not usually in publications, at

least not the ones that the PTO is used to looking at. The prior art of business methods is in the day-in-and-day-out practice of business large and small. That is hard for the PTO to find, and it has done a bad job of it.

But it is not clear that the solution to that problem is to abolish business method patents. The alternative is to change the structure of the examination process so that opportunities and incentives are created for the parties that *do* have knowledge of the relevant prior art to bring it forward. As noted above, it is unlikely that the one-click patent would have survived an appropriately designed re-examination process.

The argument that new business methods are not really inventions is, at best, an irrelevant semantic one, and, at worst, a kind of techno-snobbery that is inconsistent with how technology evolves in general. What is the substantive difference between a “tinkerer” who comes up with some new kind of machine, and a business visionary who comes up with a new method of inventory management? In either case, the invention may be made with or without the benefit of “science” in any meaningful sense. To say that one is technological and one is not is pointless. The real question, from a policy perspective, is whether the incentive provided by patent protection is necessary to bring forth the invention, and/or to protect it sufficiently to justify the investment necessary to work the kinks out and develop the raw idea into a viable commercial product or process. One can come up with individual examples of new business methods that required little development investment, but the same is true of inventions in other areas. As a general proposition, important new business methods are not dissimilar from other forms of

innovation: they often require major investments of time and money in development; there are methods other than patents (e.g. secrecy) that can sometimes be used to protect these investments, but there are also cases where, in the absence of patent protection, the risk of imitation would seriously undermine development incentives.

In summary, the problem with business patents is that many appear to have been issued for inventions that were obvious; if so, they should not have been granted. If you got rid of these invalid patents, there is not an evident problem of business method patents inordinately restricting ongoing business innovation. With the right procedures, the number of patents on business methods would probably be small, because there is a lot of prior art out there against which one would have to prove novelty and non-obviousness. But there is no fundamental economic reason why an entrepreneur who really does come up with a novel and non-obvious method of doing business needs patent protection less than an entrepreneur trying to make a go of comfortable high-heel shoes or a new way of using radio spectrum for cell phones.

Software: An open and shut case for “open source”?

Closely related to the controversy over business method patents is tremendous unhappiness over the granting of patents on software. Like business methods, this is an area where much of the prior art is not in patents, and often not in published works, but rather in practice. Again, the PTO appears not to have done a good job at making sure that applications for software patents are tested against this non-published prior art. The

result has been a deluge of patents granted on software concepts that appear not to be new.

Making things even worse, the PTO, with the apparent guidance of the CAFC, also seems to have reduced or eliminated the requirement that a patent application describe the new invention with sufficient detail to enable one skilled in the relevant art to reproduce the invention (Burt and Lemley, 2002). This drastic weakening of the “enablement” requirement seems to have led to a situation where patents can be attained on the *idea* that something could be performed with software, without the patent applicant having done much at all to implement the idea (Bessen and Hunt, 2003). The result has been a flood of patent applications on myriad diverse software ideas; in principle the recipient of such a patent then has the right to exclude others from implementing the covered software idea, despite the fact that they have never implemented, or even described implementing, the idea themselves.

Once again, the solution to these problems does not seem to be software-specific. To encourage invention and technological change, the PTO should grant patents only on novel, non-obvious software developments, and should require the applicant to describe the covered software in some detail, so that patents only go to people who have created something rather than to those who merely thought about creating something. The horror

stories about ridiculous software patents would then go away along with the PB&J, one-click, and other disasters.¹⁴

There are some who would go further and argue that patents are fundamentally inapplicable to software (Bessen and Maskin, 2001). This argument has two related pieces. The first is the straightforward observation that software innovation was flourishing *before* the 1980s, when the CAFC clarified and broadened the patentability of software. This seems to show that patents for software are not necessary. More fundamentally, it is argued that the software development is by its nature so cumulative that it is impossible to parse out the contribution of one developer sufficiently to grant patent rights, and it is counterproductive to try to do so because subsequent development will be hampered. These ideas are embodied in what has come to be called the “open source” movement, which argues that all software should be in the public domain, available for all to use, modify, and build on as they see fit.¹⁵

It is surely true that there was software innovation before software patents were widely used. As with all other technologies, it is unlikely that software development would grind to a halt without patent protection. And it is also true that software

¹⁴Burt and Lemley (2002) argue that the very logic that allows the CAFC to judge a sketchy description of what software *might* do as satisfying the enablement requirement will, inevitably, lead the court to judge many of the software patents currently being granted by the PTO to be invalid on the grounds of obviousness. Of course, it would be far preferable not to grant all those valid patents to begin with.

¹⁵See, e.g. <http://www.gnu.org/philosophy/savingeurope.html> or <http://www.freepatents.org/>.

innovation is a highly cumulative process. But the reality is that virtually all innovation is a highly cumulative process, and the patent system has been struggling with the tradeoffs that implies for a long time.¹⁶ The relevant question is: on balance, would a *properly administered* regime of software patents foster innovation, by allowing parties that make true breakthroughs a measure of protection to reduce the risks of commercializing that development? As with business methods, we haven't had a test of such a system because the PTO has failed to implement the requirements of novelty, non-obviousness and enablement.

If the overall patent system were reformed as we have proposed, the only software that would be patentable would be that which truly represents a non-obvious step forward, and the implementation of which is laid out in some detail. Granting patents of this sort would not stop others who wish to work within the open-source paradigm from doing so, and would not prevent open-source advocates from arguing their case and trying to convince computer users not to buy patented software. It may be that the advantages of open-source development are sufficient that many or most software developers would choose to forego patents and work within the open-source paradigm instead. If the PTO were doing its job properly, any software that is developed and published freely by open-source advocates or other programmers could never itself be patented, because no one could ever claim novelty in having created it. So a properly functioning patent system is not inconsistent with a vibrant open-source software

¹⁶For a discussion of historical disputes over patenting in the auto, aircraft, semiconductor and computer industries that bear much resemblance to the current software and business methods controversies, see Merges and Nelson (1990).

movement. The real enemy of open-source software—and software innovation more generally—is poor implementation of software patents, not the concept. The real question is whether a programmer who has a truly new software invention ought to have the *option* of patenting it rather than making it open-source. No one has put forth a convincing argument why that choice should not lie with the innovator, rather than being made for the entire industry as a matter of law or policy.

Should mere mortals control the human genome?

The last important area in which the CAFC expanded patentability is biotechnology. And again, there has been much concern about the granting of patents that appear obvious in light of previous developments, and which grant broad rights that seem to cover with one patent many diverse possible uses. As with software and business methods, the solution lies in rigorous enforcement of the fundamental rules governing obviousness and novelty.

An additional concern in biotechnology is the patenting of research tools. It is feared that the need to pay royalties on multiple distinct research tools in order to market a given product is or will soon retard the inventive process. But again, one must be careful to distinguish the problem of bad patents from an allegation that patents are bad. If patents are granted only on novel, non-obvious inventions, then researchers will have to pay royalties to others only for the use of research tools that were truly invented by the patent owner. Arguably, if the PTO is doing its job, a patented research tool will be one

that might not have been available at all, if the researcher who secured the patent had not developed it. It does not seem unreasonable, in such a case, for a royalty to be paid.

Does one patent “size” really fit all?

Lurking in the background of the preceding paragraphs is the over-arching question of whether we should have one set of patent rules that govern all inventions, or whether the system can be made more efficient by tailoring patent rules to the specific attributes of different technologies. In the world of theoretical patent analysis, it is easy to show that the attributes of patent protection should vary depending on the characteristics of the technology. Thus there appears to be a fundamental theoretical case for differential patents, and perhaps even for permitting patenting of some technologies but not others. The problem with using this theory as a basis for policy, however, is that the technology characteristics that could provide the basis for differential patent treatment are typically abstract and difficult to quantify empirically. It is easy, for example, to talk about cumulative innovation as a theoretical phenomenon, but it is very difficult to say concretely whether invention is more or less cumulative in one sector versus another. So while there is a theoretical case for a system that is not uniform, there is no theoretical or empirical basis for saying specifically how patent treatment should differ across specific technologies.

There is also a strong practical argument against differential treatment. Simply put, differential treatment is hard to implement, because as soon as patentees in a particular category get treatment that is different from everyone else, there will be an

inevitable tendency for people to position themselves to get the most favorable treatment. An example can be seen in the PTO's efforts to deal with the outrage over business method patents by instituting a special internal review of all business method patents, on the grounds that the prior art is difficult to identify.¹⁷ In effect, patents that fall in a particular patent class are examined twice, to try to make sure that non-traditional prior art is not missed. The result has been a decline in applications in the targeted class, but a continued rise in applications related to business methods more broadly defined. This suggests strongly that applicants have been going out of their way to classify their patents outside of the class targeted for special (more rigorous) treatment. As a result, the PTO's efforts, however well intended, are not likely to solve the broader problem of invalid business method patents being granted.

By and large, the presumption today is that everyone gets the same patent treatment.¹⁸ Without this presumption, there would be tremendous pressure by particular industries to get features in "their" patents that they found desirable. Of course, the arguments for these preferences would always be couched in public interest terms, but when an industry lobbyist starts talking about the public interest, we all know it is a good time to keep an eye on the consumer's wallet.

¹⁷<http://www.uspto.gov/web/offices/com/sol/actionplan.html>.

¹⁸Burt and Lemley (2002) argue that this uniformity is something of an illusion, because the CAFC interprets the uniform rules differently in different industries.

Even in the current system, where the general presumption is for uniformity, there are always pressures for special treatment. These pressures have been particularly acute in the pharmaceutical industry, where Congress opened the door for such gaming when it passed the Hatch-Waxman Act, which allowed the PTO to extend the length of patent coverage for drugs that had languished for an extended time in regulatory review process.¹⁹ In recent years, bills have been introduced in Congress to provide extended patent life for specific drugs, such as the allergy drug Claritin (Babcock, 1999). In the fall of 2002, a provision protecting a vaccine made by pharmaceutical maker—and large donor to Republican campaigns—Eli Lilly was inserted at the last minute into President Bush’s domestic security legislation. While Senate Majority leader Bill Frist of Tennessee had pushed such a provision earlier, he has denied inserting it in the domestic security bill. The provision became law despite the fact that *no one*, either on Capital Hill or in the White House, is willing to admit putting it in the final version of the bill (Stolberg, 2002). This kind of shenanigans would likely be much worse in a world in which it was broadly accepted that differential patents for different technologies were appropriate.

So there is grave danger in trying to “fix” the problems perceived to be associated with patents in particular areas by fooling with specific differential patent treatment for

¹⁹The rationale for the act (more formally known as the Drug Price Competition and Patent Term Restoration Act of 1984) was to ensure that each new drug had a minimum period during which it was on the market and protected by patent coverage. While the act called for these extensions to be reviewed and granted in a formal process by the PTO, the introduction of special legislation geared towards particular drugs soon followed.

these technologies. And this danger is simply not justified by evidence that the problems in business methods, software and biotechnology derive from the unique properties of these technologies. Rather, the relative inexperience of the PTO with these technologies, combined with their critical importance for the economy, has made the broader, more fundamental problems with the system most evident. It is vitally important to fix the problems with patenting in these areas—but the way to do that is to fix the problems with the patent system more generally.

A Less Kind, Less Gentle Patent System

Economic analysis does not support abolishing patents, and even weakening the fundamental presumption of validity for appropriately issued patents would carry serious policy risks. The alternative is to make sure that, before they can be used to restrict the commercial activities of competitors, patents have gotten the appropriate scrutiny to ensure their validity. At the same time, we need to accept that the PTO will still make mistakes, and create a judicial system that deals with those mistakes in a balanced way. Doing this without an infeasible increase in resources for the Patent Office will require that the system be significantly modified. The modifications should be carefully tuned to create incentives so that private parties have the incentive and opportunity to bring information to bear, but have limited incentive and opportunity to act simply to gum up the works.

Possible changes that have been proposed by various groups studying the patent system are:

- Greater resources devoted by the Patent Office to the process of examination, and the efficient use of these resources to bring the day-to-day operations of the PTO into the 21st century;
- The institution of pre-grant opposition, whereby outside parties could provide information on prior art to the examiners before a patent issues;
- The institution of effective re-examinations of granted patents, with a true opportunity to prove invalidity before an open-minded re-examiner, combined with appropriate incentives to discourage frivolous requests for re-examination; and
- Enhanced scope for judges or specially appointed masters to decide technical issues of novelty and obviousness.

The idea of these reforms is to harness the incentives of private parties to bring information to the table in an efficient way. And they respect the Rational Ignorance principle, by bringing to bear a sequence of more rigorous (and hence more expensive) investigation, as the stakes get higher. Most patents will continue to get a relatively cursory review and then be forgotten. More important ones will get a more rigorous review, and one can presume the fewer mistakes will be made in important cases as a result. For the few cases that really matter *and* the PTO still got it wrong, the courts will provide a more balanced and reliable final determination as to patent validity. As a result, the uncertainty and patent blackmail that increasingly threaten the whole innovation system should be reduced.

References

The Advisory Commission on Patent Law Reform, "A Report to the Secretary of Commerce," August 1992.

Charles Babcock, "Patent Fight Tests Drug Firm's Clout: Claritin Maker Goes All Out in Congress" *Washington Post*, (October 30, 1999) A1.

James Bessen, and Robert M. Hunt, "An Empirical Look at Software Patents," Unpublished working paper, Research on Innovation and Federal Reserve Bank of Philadelphia, 2003.

James Bessen, and Eric Maskin, "Sequential Innovation, Patents, and Imitation," Working Paper #00-01, Department of Economics, Massachusetts Institute of Technology, 2001.

Dan Burt, and Mark Lemley, "Is Patent Law Technology Specific?," *Berkeley Technology Law Journal*, 17 (2002) 1155-1206.

Ronald W. Clark, *Einstein: The Life and Times*, London: Hodder and Stoughton, 1973.

Cohen, Wesley and Stephen Merrill, Eds., Patents in the Knowledge-Based Economy, National Academy Press, www.nap.edu/catalog/10770.html (2003)

Gallini, Nancy and Suzanne Scotchmer "Intellectual Property: When Is It the Best Incentive System?," *Innovation Policy and the Economy*, 2 (2002) 51-77.

Stuart J.H. Graham, Bronwyn Hall, Dietmar Harhoff, and David C. Mowery, 2002, "Post-Issue Patent 'Quality Control': A Comparative Study of US Patent Re-examinations and European Patent Oppositions," Working Paper No. 8807, National Bureau of Economic Research, 2002.

Hall, Bronwyn H., Stuart J. H. Graham, Dietmar Harhoff, and David C. Mowery "Prospects for Improving U.S. Patent Quality via Post-grant Opposition," *Innovation Policy and the Economy*, 4 (2004) __-__.

Intellectual Property Owners Association, Statement Submitted to the Subcommittee On Courts, the Internet, and Intellectual Property, Committee on the Judiciary, U.S. House Of Representatives, Hearing on the "U.S. Patent and Trademark Fee Modernization Act Of 2003," April 3, 2003.

Adam B. Jaffe, and Josh Lerner, *Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What To Do About It*, Princeton: Princeton University Press, 2004.

Michael K. Kirk, Statement before the Subcommittee On Courts, the Internet, and Intellectual Property, Committee on the Judiciary, U.S. House Of Representatives, Hearing on the “United States Patent and Trademark Office Fee Modernization Act Of 2003,” April 3, 2003.

Samuel Kortum and Josh Lerner, “Stronger Protection or Technological Revolution: What is Behind the Recent Surge in Patenting?,” *Carnegie-Rochester Conference Series on Public Policy*, 48 (1998) 247-304.

Samuel Kortum and Josh Lerner, “Unraveling the Patent Paradox,” Unpublished working paper, University of Minnesota and Harvard University, 2003.

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

Robert P. Merges and Richard R. Nelson, “On the Complex Economics of Patent Scope,” *Columbia Law Review*, 90 (1990) 839-916.

Merrill, Stephen A., Richard C. Levin and Mark B. Myers, Eds., A Patent System for the 21st Century (Report of the Committee on Intellectual Property Rights in the Knowledge-Based Economy), Board on Science Technology and Economic Policy, National Research Council (“STEP Report”), www.nap.edu/html/patentsystem (2004)

“Patents: Improving Quality and Curing Defects,” Hearing before the Subcommittee on Courts, the Internet and Intellectual Property, Committee on the Judiciary of the House of Representatives, May 10, 2001.

Ronald J. Stern, Statement Submitted to the Subcommittee on Courts, the Internet and Intellectual Property, Committee on the Judiciary, U.S. House of Representatives, on The U.S. Patent and Trademark Fee Modernization Act of 2003, April 3, 2003.

Sheryl Gay Stolberg, “A Capitol Hill Mystery: Who Aided Drug Maker?,” *New York Times*, (November 29, 2002) A35.

U.S. Federal Trade Commission, To Promote Innovation, The Proper Balance of Competition and Patent Law and Policy, www.ftc.gov/os/2003/10/innovationrpt.pdf (2003).

Figure 1: Annual Patent Applications and Awards

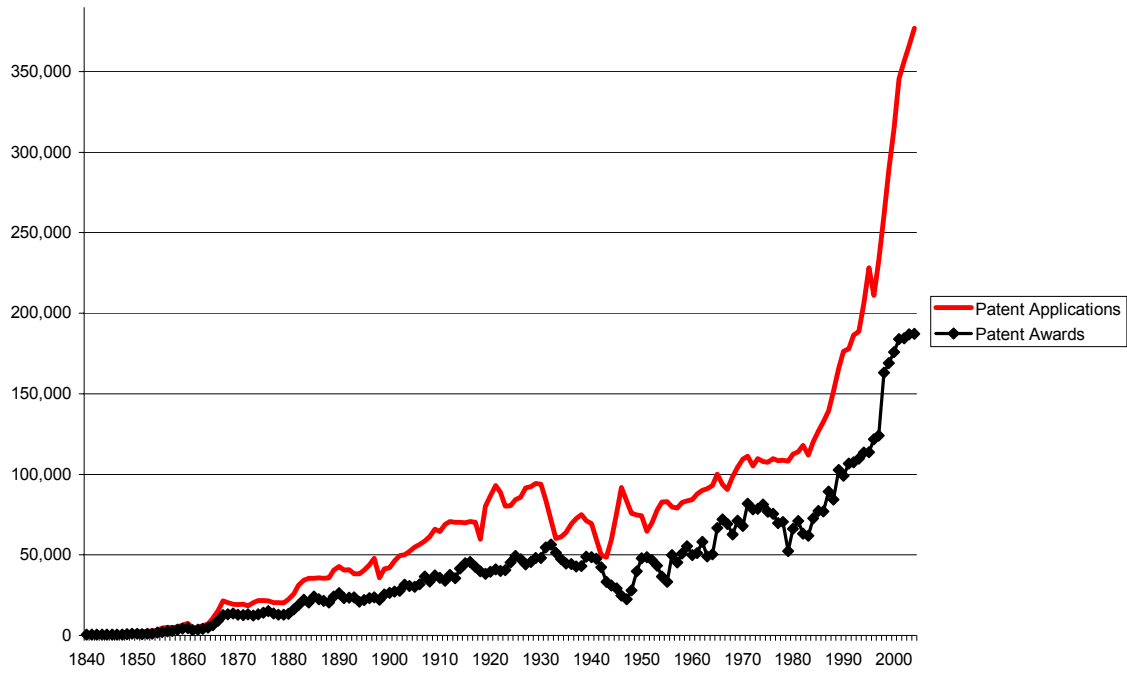


Figure 2: Number of Patent Suits Initiated

