Volume Title: Innovation Policy and the Economy, Volume 8

Volume Author/Editor: Adam B. Jaffe, Josh Lerner and Scott Stern, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-39121-3

Volume URL: http://www.nber.org/books/jaff08-1

Conference Date: April 12, 2007

Publication Date: April 2008

Chapter Title: Patent Reform: Aligning Reward and Contribution

Chapter Author: Carl Shapiro

Chapter URL: http://www.nber.org/chapters/c5303

Chapter pages in book: (111 - 156)

4

Patent Reform: Aligning Reward and Contribution

Carl Shapiro, *University of California at Berkeley*

Executive Summary

Economists and policymakers have long recognized that innovators must be able to appropriate a reasonable portion of the social benefits of their innovations if innovation is to be suitably rewarded and encouraged. However, this paper identifies a number of specific fact patterns under which the current U.S. patent system allows patent holders to capture private rewards that *exceed* their social contributions. Such excessive patentee rewards are socially costly as they raise the deadweight loss associated with the patent system and discourage innovation by others. Economic efficiency is promoted if rewards to patent holders are aligned with and do not exceed their social contributions. This paper analyzes two major reforms to the patent system designed to spur innovation by better aligning the rewards and contributions of patent holders: establishing an independent invention defense in patent infringement cases, and strengthening the procedures by which patents are reexamined after they are issued. Three additional reforms relating to patent litigation are also studied: limiting the use of injunctions, clarifying the way in which "reasonable royalties" are calculated, and narrowing the definition of "willful infringement."

I. Introduction

A growing chorus of scholars and practitioners are expressing concerns about the operation of the U.S. patent system.[1] While there is no doubt that the U.S. economy remains highly innovative, and there is no doubt that the patent system *taken as a whole* plays an important role in spurring innovation, the general consensus is that the U.S. patent system is out of balance and can be substantially improved. Problems with the system seem to reside especially in the fields of information technology and biotechnology.

Going back to economic first principles, this paper argues that the efficiency of the current U.S. patent system can be significantly enhanced by

reforming the system in two major ways. First, an independent invention defense could be established in patent infringement cases. Under such a defense, a party accused of patent infringement can avoid liability if it can establish that it independently invented the patented technology. Second, a much more effective procedure could be established for reexamining selected patents after they are issued by the U.S. Patent and Trademark Office (PTO). The central goal of this paper is to illuminate the strong economic logic behind patent reform proposals along these lines, without delving deeply into the specifics of how they would be implemented in practice. Additional reforms are also addressed.

The core problem with the current U.S. patent system explored here is that, in certain identifiable circumstances, the patent system predictably provides excessive rewards to patent holders. The term "excessive rewards" is defined here to mean rewards that exceed the patentee's actual contribution to economic welfare. Excessive rewards are not benign as they come at the expense of other parties. In the short term, excessive rewards cause deadweight loss by raising the cost of goods and services supplied by technology users who pay excessive royalties or incur other costs to avoid liability for patent infringement. Even more important, in the long term, excessive rewards to some patentees reduce economic efficiency by discouraging innovation by other parties whose costs are elevated or whose options are restricted.

This central problem has multiple manifestations, which are reflected in a number of seemingly disparate criticisms commonly made of the system, including the following:

- Patents are regularly issued covering technologies that should, in fact, be considered "obvious."[2]

- Patents are commonly issued by the PTO for technologies that are not, in fact, "novel," due in part to the difficulties of finding prior art during the PTO examination process.[3]

- Patent applicants often obtain patents with overly broad claims, including claims explicitly drafted in continuation applications to capture products introduced into the market after the patent application was initially filed.[4]

- The owner of a patent covering one feature in a complex product can use the threat of an injunction as a powerful weapon when bargaining over royalties with parties accused of infringement.[5]

- Standard setting organizations face the continual threat of "patent ambush" by owners of patents that are essential to the standards they adopt.[6]
- The way in which "willful" infringement is defined in practice gives patent holders additional bargaining power with alleged infringers who did not copy the patented technology.[7]
- The concept of "reasonable royalties," which is used to assess damages in some patent infringement cases, tends to yield excessive royalties in cases where the patent covers one minor feature in a complex product.[8]

Those who defend the current system, or simply are cautious about reforming it, are naturally and understandably concerned that reducing the rewards to patent holders will retard innovation, wounding the goose that lays the golden eggs of economic growth. The burden of the argument, therefore, rests on those proposing reforms to show that any proposed changes will promote overall long-run economic efficiency, taking full account of their effects on innovation.

The fundamental building block for the arguments made in the following is that economic efficiency is promoted when the rewards provided to patent holders are aligned with their actual social contributions. Efficiency is not a monotonic function of the rewards provided to patent holders: excessive rewards, just like inadequate rewards, can reduce efficiency and stifle innovation.[9] The patent reform proposals considered here attempt to rectify situations in which rewards to patent holders are likely to *exceed* their social contributions. Reducing such excessive rewards promotes economic efficiency and encourages socially valuable innovation.

Importantly, the two primary reform proposals considered here, expanding the independent invention defense and the use of reexaminations, would have highly targeted effects on specific types of patent holders. They would *not* cause an across-the-board reduction in the rewards to patent holders generally. In this respect, they are quite different from the classic instrument of patent policy traditionally studied by economists, namely patent length. In theory, determining the optimal patent length requires an understanding of the relationship between the rewards provided to patent holders and the extent of inventive activity, that is, that one has an estimate of the elasticity of supply of inventions. Estimating this elasticity is notoriously difficult. Moreover, there is enormous variation in the economic significance of different patents and in

the costs and risks associated with discovering and developing different technologies. Because the proposals studied here have targeted effects and only seek to bring the reward to patent holders closer in line with their social contributions, their attractiveness does not hinge on the elasticity of supply of inventions.[10] If a patent holder's reward exceeds its social contribution, reducing its reward promotes economic efficiency, regardless of that elasticity.

Nor do the proposals considered in the following rely on the ability of the PTO or the courts to distinguish among inventions based on their benefits, cost, or risk. Rather, they are designed systematically and automatically to reduce the exclusionary powers of patent holders in cases where patent holders' rewards predictably exceed their social contributions. These reforms are feasible within the context of a unitary patent system (i.e., one that does not vary by field of technology), recognizing the very limited information about benefits, costs, and risks available to the PTO and the courts. The welfare test for the proposals studied here is whether they promote economic efficiency regardless of the distribution of benefits, costs, and risk across patented inventions. In the short run, that is, after the invention has been achieved, reforms that reduce excessive rewards improve ex post efficiency. More important, in the long run, to the extent that such reforms affect research and development (R&D) and patenting decisions, efficiency is further improved.

Section II discusses, motivates, and elucidates the general concept of aligning private rewards with actual social contribution using patents. A number of settings are identified in which patentees' rewards tend to exceed their social contributions. These general ideas are then applied in the subsequent sections. Section III studies the impact of greatly expanding the independent invention defense in patent infringement cases. Section IV considers the effects of enhancing the role of patent reexamination. Section V discusses the treatment of injunctions, the determination of "reasonable royalties," and the doctrine of willful infringement in patent infringement cases. Section VI concludes.

II. Aligning Patentee Rewards and Contributions

A natural starting point in thinking about the relationship between rewards and contributions using the patent system is to ask whether a patent holder's reward should *equal* the social contribution resulting from its invention. In other words, is "full appropriation" by the inventor of the social benefits resulting from its invention optimal? To explore

this question, we define the *appropriability ratio* as the ratio of the patent holder's payoff to the social contribution associated with the patented invention. Should the appropriability ratio equal unity?

The idea that full appropriation is optimal is rooted in the basic economic concept of externalities: the activity of conducting R&D will be undersupplied if it generates positive externalities and oversupplied if it creates negative externalities. According to this reasoning, if an inventor can appropriate only a fraction of the social value generated by its invention, his or her R&D activities will be undersupplied. Of course, the central goal of the patent system is to increase appropriability to spur innovation. But this certainly does not imply that full appropriation is optimal, even if it is feasible. Plus, while stronger patent rights may increase the returns for some innovators, they will decrease the returns for others, especially those who do not rely on patents.

Perhaps due to its simplicity and intuitive appeal, the idea that the rewards to patent holders are too low if they fall short of the social contributions associated with their inventions has influenced innovation policy. Combined with the empirical evidence that many inventions generate positive externalities, generally known as spillovers, the full appropriation intuition suggests that the patent system generally underrewards innovators.[11] If one accepts this reasoning, reforms that reduce the rewards to patent holders seem ill advised.

Full appropriation can indeed be shown to be optimal in a very specific setting with several very strong assumptions: if only a single firm is capable of pursuing the invention at issue, if the firm can be rewarded in a way that does not cause any inefficiency, for example, through a monetary prize funded by nondistortionary taxes, and if the firm's contribution can be accurately observed by the authority awarding the prize. The appendix contains two simple models along these lines. However, the reasoning behind this full appropriability result is not robust, and it certainly is not a reliable guide to patent policy for several very important reasons that we now discuss.

A. Multiple Firms Pursuing the Same Invention

The full appropriability result, as usually conceived, breaks down rather dramatically once one considers inventions that can be pursued by multiple firms. When multiple firms can pursue a given invention, granting a patent to the first successful firm and setting the patentee's reward equal to the social contribution associated with the *invention* leads to

wasteful duplication of effort, dissipating the social benefits of the invention. This is true even if the patent monopoly does not cause any ex post deadweight loss.[12]

In their model of patent races, Dasgupta and Stiglitz (1980) show that full appropriation causes a "common pool" problem, leading to excessive R&D expenditures that dissipate the social benefits of the invention. Tandon (1983) obtains a similar result in a model in which multiple firms make risky R&D investments. Tandon shows that the optimal degree of appropriability is smaller, the lower is the cost of pursuing the invention. For "nearly obvious" inventions, that is, inventions that can be achieved with high probability at modest cost, the optimal appropriability ratio is low, even before we account for the positive relationship between appropriability and deadweight loss. The robust lesson from this line of research is that the optimal appropriation ratio is less than unity under a conventional patent system in which the first firm to achieve the invention receives a reward in the form of exclusive rights.[13]

Why does the basic "externality" intuition fail so badly when multiple firms are pursuing the same invention? What is wrong with rewarding the full social contribution associated with the invention to the first inventor? Answering this question requires that one distinguish between the social contribution associated with the *availability of the invention* and the social contribution of any given *inventor.* If only one party can pursue the invention, this distinction evaporates, and the full appropriation intuition is valid. However, if two or more parties can achieve the invention independently, this distinction becomes very substantial indeed.

Consider the polar case in which two parties achieve the same invention at precisely the same time. Under a conventional patent system, the social contribution associated with the invention is the same as if just one party had achieved the invention at that time: the invention is available to society but controlled by a single party. Strikingly, each inventor's social contribution in this case is nil: the invention would have been available to society even if that inventor had not discovered the invention.[14] Any positive award to either inventor will exceed that party's incremental social contribution. Taking seriously the notion of rewarding inventors based on their incremental contributions would suggest a very different system, under which the invention is placed in the public domain once it has been discovered independently by two or more parties. Such a system may be unfamiliar, and even appear strange, but it would, in fact, be faithful to the idea of aligning each inventor's rewards with its incremental social contribution.[15]

No, this paper is not actually proposing replacing the current patent system with one that awards a patent to the first inventor, with the life of the patent extending only until the date of subsequent invention, at which time the invention is placed in the public domain. The goal here is more modest: to establish that allowing the first inventor to appropriate a significant share of the social benefits associated with its invention can greatly overreward the first inventor in situations where subsequent independent invention occurs. Using the standard definition of the appropriability ratio given in the preceding, the optimal appropriability ratio with independent invention can be very low. This reflects the fact that the contribution of any one *inventor* is a small fraction of the contribution associated with the *invention*.

This powerful economic point will be central in section III when we consider the economic effects of expanding the independent invention defense: such a defense only arises in the event that two or more parties independently discover the same invention. Multiple independent invention is most likely in situations where the underlying knowledge base in the public domain is advancing rapidly, so many incremental improvements are "in the air." It is no coincidence that this appears to be common today in the information technology and biotechnology sectors, the very sectors where the call for patent reform is loudest. Multiple independent discovery also is most likely for inventions with a high benefit to cost ratio, which includes inventions that are easily achieved, that is, close to the boundary of being "obvious."

B. The Public Good Aspect of Patent Challenges

We now consider the relationship between reward and contribution for a probabilistic patent, that is, a patent that may be invalid because the patent holder did *not*, in fact, contribute a novel, nonobvious invention to society. As emphasized by Lemley and Shapiro (2005) among others, patents are invariably probabilistic when they are issued, and there is considerable evidence that a large number of issued patents are weak. Following Farrell and Shapiro (2008), we identify circumstances in which the owner of a probabilistic patent can obtain a reward that exceeds its social contribution. Accounting for the fact that the PTO often fails to find relevant prior art, a class of situations can be identified in which patentees' rewards exceed their contributions.

What is the social contribution of a party that obtains a probabilistic patent, that is, a patent that might be invalidated if a more thorough

search for prior art is conducted through reexamination or litigation? Farrell and Shapiro (2008) show conceptually how to measure the contribution associated with a probabilistic patent that is licensed and not tested by litigation or reexamination. As usual, social contribution is the difference between realized total welfare and the welfare that would have resulted in the absence of the patent holder. The key is to recognize that, if not for the patent holder, two outcomes might have arisen. If the patent holder truly *did* cause the patented technology to be available to society, that is, if the patented technology is novel and nonobvious, then without the patent holder the patented technology would not be available to society.[16] Alternatively, if the patented technology is not novel, or is obvious, then without the patent holder the patented technology would be available *and in the public domain*. The former case corresponds to the situation normally studied, in which the patent is surely valid, and the patent holder's social contribution is clearly positive. In the latter case, however, the patent holder's social contribution is actually *negative*, namely the deadweight loss associated with the patent monopoly: while inventing nothing, the patent holder has obtained exclusive rights to a technology that would otherwise have been in the public domain.

Applying this conceptual framework, Farrell and Shapiro (2008) study in detail the case of a patent holder who licenses its probabilistic patent to a downstream oligopolistic industry. They consider a patented technology that reduces the marginal cost of producing downstream products by v, the "value" of the patented technology.[17] They show that the patent holder's profits exceed its social contribution if the running royalty rate exceeds the expected cost saving attributable to the patent holder, that is, the value of the patented technology discounted by the patent strength.[18] For example, if the patented technology reduces costs by $10 per unit, and if the patent is valid with 30 percent probability, then the patent holder's profits exceed its contribution if the running royalty rate exceeds $3 per unit. The appendix derives this intuitive benchmark in the important special case where the downstream market is perfectly competitive.

Farrell and Shapiro (2008) then identify circumstances under which owners of weak patents can indeed obtain per-unit royalties in excess of the benchmark level.[19] This is not possible for ironclad patents: if the owner of an ironclad patent tried to charge a running royalty greater than the value of the patented technology, potential licensees would just say no and refrain from using the patented technology. Likewise, potential customers would purchase substitute, noninfringing products.

But this logic does not extend to probabilistic patents. How will a potential licensee respond if the patented technology reduces costs by $10 per unit, if the patent is valid with 30 percent probability, and if the patent holder offers to license the patent for, say, $6 per unit, twice the $3 benchmark level? The potential licensee faced with such an offer will not simply refrain from using the patented technology: better to pay a royalty of $6 per unit than forego the $10 per unit value of the patented technology. The key question, then, is whether the potential licensee will find it more profitable to pay $6 per unit or to infringe and face the prospect of being sued for infringement.

What are the benefits and costs of licensing versus litigation from the perspective of a single technology user? Under U.S. patent law, any one technology user who challenges the patent is providing a public good for other technology users: if the patent is invalidated, other technology users benefit because they, too, can then use the (formerly) patented technology freely.[20] As emphasized by Farrell and Merges (2004) and Miller (2004), this means that any one firm challenging a patent is providing a public good to other technology users and to consumers who purchase products that use the patented technology. As usual, the market undersupplies public goods, in this case patent challenges.

Farrell and Shapiro (2008) show how owners of weak patents can exploit the public good nature of patent challenges to obtain royalties in excess of their social contributions. These excess rewards occur if the patent holder is licensing to multiple downstream firms that compete against each other. In this situation, the owner of a weak patent can obtain a surprising high per-unit royalty rate in the licensing equilibrium. The key driver here, emphasized in Farrell and Merges (2004), is "relatively." Because the downstream firms compete against each other, the profits of any one downstream firm are sensitive to that firm's costs *relative* to the others. In the licensing equilibrium, a single firm that challenges the patent and *loses* places itself at a competitive disadvantage relative to its rivals who signed licenses. But a single downstream firm that challenges the patent and *wins* does not enjoy a competitive advantage relative to its rivals who signed licenses, as in that event *all* downstream firms are relieved of the duty to pay royalties.[21] These results are derived in a model with zero litigation costs; they become even stronger if litigation costs give additional bargaining power to patent holders.[22]

This analysis will be highly relevant in section IV, where we consider the benefits of improving and expanding patent reexaminations. In fact, Farrell and Shapiro (2008) derive expressions for the benefits of enhanced

reexamination prior to licensing.[23] One of their findings is that these benefits can be large for patents of intermediate strength that are licensed to multiple downstream rivals, but small (or even negative) for patents licensed to downstream firms that do not compete against each other.[24]

C. *Specific Investments and Patent Hold-Up*

Another important situation in which patent holders may be able to capture more than their social contribution arises if they can act opportunistically, appropriating a portion of the investments made by others that are specific to the patented technology. Such opportunism can easily arise when technology users make such specific investments before they negotiate the terms of a patent license, perhaps because they are unaware of the (pending or actual) patent at the time these investments are made, or because the patent is broad and vague, or because of other obstacles to negotiating and signing licenses at such an early stage in product development.

The economics of opportunism are well understood, and there is nothing at all exceptional about applying these ideas to patent licensing.[25] To illustrate the basic idea, consider a manufacturer that is designing its product and the holder of a patent covering technology that adds value to the product. Suppose the manufacturer is considering two product designs. Design A uses the patented technology, if the product is made using Design A, it will command a price of $120. Design B does not use the patented technology; if the product is made using Design B, it will command a price of only $100. All production costs are the same for both designs. Thus, the contribution of the patented technology is $20, namely the value it adds to the product.

Suppose that the manufacturer must incur certain costs to design its product. Either of the two designs requires the manufacturer to incur $30 in costs that are common to the two designs. Each design also involves $40 in design-specific costs. The patent holder and the manufacturer bargain over royalties; for simplicity, suppose that they have equal bargaining skill, so they split equally any gains from trade. Suppose that the patent is ironclad, and the manufacturer is prohibited from selling products using Design A if it lacks a license to the patented technology.

If the two parties bargain before the manufacturer has incurred any design costs, the negotiated royalty will equal $10 as the patented technology is worth $20, and these gains from trade are split equally. The appropriability ratio is 50 percent ($10/$20). However, if the two parties

bargain *after* the manufacturer has incurred the costs associated with Design A, the negotiated royalty will equal $30. Why? The gains from trade in this situation are $60: the $20 in inherent value associated with the patented technology plus the $40 of design-specific costs that will be wasted if the two parties cannot agree on licensing terms and the manufacturer is forced to revert to Design B. Intuitively, the manufacturer is in a weak bargaining position after it incurs $40 in costs specific to Design A. The appropriability ratio is now 150 percent ($30/$20).[26] As a general proposition, if a technology user negotiates over royalties with a patent owner after the user has incurred costs that are specific to the patented technology, the reward to the patent holder can exceed its contribution. In practice, a key question is whether the patent system provides a mechanism whereby technology users can avoid placing themselves in this situation. For an ironclad patent, this turns on whether the manufacturer is aware of the patent at the time it designs its product. Such awareness may be impossible if the patent application has not yet been published and very difficult for questionable patents and for patents with broad and vague claims. This point will be very important in section III, when we consider an independent invention defense for those who practice the invention before it is disclosed by the patentee.

Concerns about patent holdup are by no means simply theoretical. They are very real in many patent infringement cases, where licensing takes place in the shadow of litigation that may lead to an injunction preventing the manufacturer from selling products that have an infringing component. Concerns about patent holdup are sometimes expressed colorfully in terms of so-called patent trolls, who allegedly extract excessive royalties by threatening to obtain injunctions preventing infringing firms from selling complex products to which the patented technology contributes only a small fraction of the value.[27]

Shapiro (2006a) extends the analysis of patent holdup to probabilistic patents. He quantifies the extent to which the rewards to owners of probabilistic patents exceed the no-holdup benchmark when these patent owners can threaten to obtain permanent injunctions.[28] The impact of patent holdup is most pronounced if the patented technology covers a relatively small component of a complex product containing many other innovative features and where redesigning the product to avoid infringing would involve significant costs or time. Patent holdup problems also depend upon patent strength, that is, the probability that the patent will be held valid and infringed if litigated. Returns above the benchmark level are especially pronounced for weak

patents. In fact, the owner of a weak patent can benefit from holdup even if the manufacturer is aware of it prior to making any product design decisions or specific investments. This analysis will be central in section V.A, when we consider the rules governing permanent injunctions.

The patent holdup problem is one of timing: technology users must make product design and decisions involving specific investments either (1) before they are aware of a patent that may be asserted against them or (2) before they can resolve whether the patent is valid or whether their product would indeed infringe the patent (perhaps because the claims are broad and vague). These problems can be reduced or eliminated if the patent's validity and scope can be determined sooner rather than later. This observation will be directly relevant in section IV, when we consider the benefits of improving patent reexamination procedures. Such procedures hold out the promise of providing substantial information about patent validity and scope sooner, and at less expense, than is possible through patent litigation. Critically, they also allow potential infringers, rather than patent holders, to influence or control the timing by which the patent is tested more carefully than it was during the initial examination.

D. Multiple Firms Pursuing Complementary Innovations

We now consider the common circumstance in which a number of firms are pursuing complementary innovations. Our central point here is that rewarding any one innovator with the full incremental value of its invention typically is not desirable because doing so will reduce the reward available to other complementary innovators.

To illustrate the basic economic issues that arise with complementary innovations, consider two parties who are pursuing innovations that are technical complements.[29] Suppose that the social value created if both inventions are discovered is V_B, the social value created if just invention #1 is discovered is V_1, and likewise for invention #2 and V_2. The concept of technical complementarity is captured by the inequality $V_B > V_1 + V_2$. Write $S \equiv V_B - (V_1 + V_2)$ for the "synergy" between the two inventions, that is, the added value from combining them.

Examples of complementary innovations abound, especially in the information technology sector, where multiple innovations are routinely combined to produce a single product or system. For example, if invention #1 is a faster microprocessor, and invention #2 is improved power management, together, these two innovations complement each other in improving the performance of laptop computers. Improvements in

wireless Internet access would further complement these two innovations. As a numerical example, suppose invention #1 alone adds value V_1 = \$50 to a laptop, invention #2 alone adds value V_2 = \$10, and together they add value V_B = \$80. The amount of the synergy is, therefore, S = \$80 – (\$50 + \$10) = \$20.

As is well known in other contexts, in the presence of complements, the sum of incremental contributions exceeds the total contribution. Here, if both inventions are discovered, the incremental contribution of invention #1 is $V_B - V_2$, and the incremental contribution of invention #2 is $V_B - V_1$. In our numerical example, the incremental value of invention #1, given success with invention #2, is $V_B - V_2$ = \$80 – \$10 = \$70, and the incremental value of invention #2, given success with invention #1, is $V_B - V_1$ = \$80 – \$50 = \$30. In general, the sum of the two incremental contributions is $2V_B - (V_1 + V_2)$, which equals $V_B + S$. In our numerical example, the sum of the incremental values is \$70 + \$30 = \$100, which exceeds the total value of \$80 by the amount of the synergy, namely \$20. In general, it is simply not feasible for each patent holder to receive its full incremental contribution unless the firms are subsidized by the amount S. We assume no such subsidies are available.

Under a patent system (as opposed to a system of public subsidies for R&D), in the presence of complementary innovations, there is a very real danger that establishing patent rights that give a greater return to one innovator will reduce the return to another, complementary innovator, and thus reduce the likelihood that the synergies will, in fact, be achieved (because they require two or more innovations to be made successfully). If the rewards are not spread properly across the different innovators, providing a greater return to one innovator can easily reduce overall innovation and impair long-run economic efficiency. This danger is especially great if one innovator is in a position to capture its full incremental contribution or even more, perhaps, because it has a first-mover advantage or can engage in opportunistic behavior.

As shown in the appendix, the optimal way to divide up the amount of the synergy between the two innovators, in the event both succeed, depends upon the elasticity of one innovator's probabilities of successful invention, with respect to the portion of the synergy it captures, relative to the other's elasticity.[30] These elasticities typically will be very difficult to observe, although we do know that the fraction of the synergy allocated to a firm should be very small if that firm can achieve its invention with high probability at low cost, which tends to correspond to the notion of an "obvious" or nearly obvious technology. The central

point here is that two or more innovators cannot both receive their incremental contributions; synergies must be shared, which implies partial appropriability for each. If many parties must make investments that provide complementary inputs, the resulting synergies may need to be shared among many parties.[31]

This analysis should serve as a stark reminder that the benchmark of full appropriability is simply not feasible in practice when multiple, complementary innovations are involved. Policies that spread out the available rewards among parties contributing complementary innovations are generally desirable. Arguments by any one patent holder that it should receive its full incremental contribution are incomplete in the presence of multiple, complementary innovations. Any system under which one or a few parties can appropriate a large share of these synergies is likely to be inefficient, leading to less innovation than would be possible under a more balanced system. This is especially true if the party capturing a large share of the synergies does not have a high elasticity of success with respect to its share of the synergies or if that party captures more than its incremental value, perhaps for one of the reasons discussed in the preceding.

In applying these ideas, it is important to recognize that innovative activities go far beyond the stage of research and invention, and many types of innovation do not involve patents at all. Significant risky investments may be required to develop the resulting invention into a workable product, to refine and test different designs, to assemble various complementary assets, and to make investments in manufacturing and marketing that are specific to the resulting product. All of these complementary activities are part of the process of innovation, and all may be necessary before an invention can lead to economic benefits.

The discussion here has been cast in terms of innovations that complement each other but do not build upon each other. Very similar issues arise in the equally important context of *cumulative innovation*. Cumulative innovations also are complementary but follow a somewhat different logic than used in the preceding. In the pure case of cumulative innovation, innovation #2 cannot be discovered unless and until innovation #1 is discovered. The incremental value of innovation #1 is V_1 without innovation #2 and V_B with innovation #2, much as it was in the preceding if $V_2 = 0$. But innovation #2 cannot even be contemplated unless and until innovation #1 has been achieved, which is different from the situation treated in the preceding (although similar to that case with

$V_2 = 0$). Once innovation #1 has been achieved, the incremental value of innovation #2 is $V_B - V_1$. The sum of the incremental contributions in the event that both innovations are successful is $V_B + (V_B - V_1)$, which exceeds the total value, V_B by the amount of the synergy, $S = V_B - V_1$. As in the case of complementary innovations that do not build upon each other, a patent system without subsidies cannot reward each innovator with its incremental contribution if both succeed.

There is a substantial literature on optimal reward systems with cumulative innovation. Scotchmer (2004) provides a valuable discussion of the tricky problems that arise in this context; she focuses on the role of patent breadth and the required inventive step.[32] One idea in this literature, going back at least to Kitch (1977), is that well-defined property rights combined with full appropriation can lead to efficient innovation incentives. However, this argument relies on some very strong and unrealistic assumptions: that the various potential innovators can sign contracts before they incur the costs associated with innovating and that the transaction costs associated with such contracts are minimal.[33] These conditions are very unlikely to be met in cases of cumulative innovation or in situations where many parties are engaged in simultaneous complementary innovative activities, especially where these activities involve considerable uncertainty about technical outcomes and where patents are probabilistic and contain broad and vague claims. The more recent literature has emphasized the need to spread out the available rewards among multiple innovators; see, especially, Hopenhayn, Llobet, and Mitchell (2006). Bessen and Maskin (2006) establish conditions under which patents actually retard cumulative innovation.

The idea that full appropriation by one innovator may come at the expense of other innovators, and thus is not optimal in the presence of complementary innovations, is generally relevant to any discussion of the optimal appropriation ratio. This idea will be especially important in section V, when we discuss injunctions, the determination of reasonable royalties, and the doctrine of willful infringement.

E. Deadweight Loss from Patent Monopoly

The full appropriability result breaks down for another important reason when applied to the patent system as opposed to a prize system. As shown in the appendix, the patent holder's reward should generally be *less* than its social contribution if the act of giving the reward generates

deadweight loss. This will typically be the case when the reward comes in the form of a patent, which grants exclusionary rights, rather than in the form of a monetary prize.[34] So long as a patent causes deadweight loss, the optimal appropriation ratio is less than unity. While we do not rely directly on this result in the following when we consider specific reforms to the patent system, it is well to bear this point in mind when discussing private rewards using the patent system.

F. Summary

The idea that innovation is best encouraged by equating the rewards to patent holders with their social contributions is alluring and often correct. But it is critical not to confuse this idea with the sharply different, and *incorrect*, idea that patent holders' rewards should be equated with the social contributions of the technologies that they patent. Such rewards are often excessive.

When multiple firms independently achieve the same or similar inventions, each firm's social contribution is far less than the social contribution associated with the availability of the patented technology. For probabilistic patents, the patent holder's social contribution is only equal to the social contribution associated with the patented technology in the event that the patent is valid, that is, in the event that the patented technology truly is novel and nonobvious. Plus, when multiple firms pursue complementary inventions, as frequently occurs in the information technology and biotechnology fields and is the norm in the context of cumulative innovation, it is simply not feasible, using a decentralized patent system, to reward each firm with the incremental contribution associated with the technology that it patents. Under the optimal patent reward system, all innovators appropriate their social contribution partially, not fully.

For all of these reasons, a patent system that allows patent holders to fully appropriate the social benefits associated with the technologies that they patent will tend to overreward patent holders and thus *retard* innovation in comparison with a more carefully designed system that reduces or eliminates such excessive rewards. We now consider a number of reforms to the patent system that have a targeted incidence, systematically reducing *excessive* rewards to patent holders, not *all* rewards to patent holders. These reforms are attractive regardless of the elasticity of supply of inventions.

III. Independent Invention Defense

Under U.S. patent law, if two or more parties achieve the same invention at nearly the same time, the first party to invent typically is awarded the patent and has the right to sue the other party for patent infringement if that party practices the invention.[35] Disputes over priority are handled through interference actions at the PTO.[36] Once the patent is assigned, a party accused of infringement cannot avoid liability by establishing that it independently invented the technology covered by the patent. There is no independent invention defense. Under a full-fledged independent invention defense, the first inventor would still receive a patent and could sue anyone who is practicing the invention for patent infringement, but any party sued for infringement would have the right to use the patented technology freely if it could establish that it independently invented the same technology. Would introducing such a defense promote economic efficiency?

Introducing an independent invention defense into U.S. patent law would be a very significant change, precisely because so many patented technologies are, in fact, discovered independently by more than one party. Historically, many pioneering inventions were discovered at nearly the same time by more than one party, including the telegraph (Morse and Alter), the light bulb (Edison and Swan), the telephone (Bell and Gray), and the integrated circuit (Kilby and Noyce).[37] Independent invention may be even more common for incremental innovations, especially in rapidly advancing fields such as information technology and biotechnology, where many applied ideas flowing from basic research are in the air at any given time. Plus, the more narrowly the notion of "obviousness" is defined, the easier it is for one party to obtain a patent on an incremental innovation that other parties are achieving at the same time.

To fix ideas, consider the following basic fact pattern. Party A achieves an invention first and files for a patent. Party B achieves the invention second, but entirely independent of Party A. Under U.S. patent law, the patent is awarded to Party A, assuming its invention otherwise qualifies for patentability. Armed with its patent, Party A can prevent Party B from practicing the invention. Independent second inventors come up empty handed. With an independent invention defense, Party A would still obtain the patent, but Party B would have the right to use the patented technology free of charge.

A. *Efficiency Effects*

How would an independent invention defense affect economic efficiency? In particular, how would the defense affect the alignment between reward and contribution?

Begin with Party A, the first inventor. Based on the discussion in section II.A about multiple independent invention, it should be clear that the defense would generally bring Party A's reward closer in line with its contribution. Party A's contribution includes achieving the invention before Party B; with the defense in place, Party A would still be rewarded with exclusive rights during this interim period. Party A's social contribution clearly does not include Party B's use of the invention. Because Party A retains the patent and can exclude others, that is, the invention is not placed in the public domain after multiple independent discovery, there are good reasons to believe that Party A's rewards will remain excessive even with the defense in place. Offsetting this are possible spillovers attributable to Party A, such as follow-on noninfringing innovations that are spurred by Party A's patent disclosure. However, many of these disclosure spillovers, as well as other spillovers arising from reverse engineering or the leakage of know-how, may not be attributable to Party A; with the defense in place, they would also result from Party B practicing the invention. Unless the spillovers *specific to Party A* are quite large, Party A's reward will still be excessive even with the defense in place as Party A retains the ability to sue others for patent infringement even after Party B's independent invention.

What about Party B? With the defense in place, Party B makes a real social contribution by converting the patent monopoly into a patent duopoly. Party B's rewards are increased by its ability to invoke the defense, but not to the point that they are excessive. Party B simply has the right to use the invention it discovered independently, in competition with Party A. To the extent that there are spillovers specific to Party B's invention, raising Party B's reward moves it closer to Party B's contribution, without granting Party B any exclusionary rights. Such a shift must have a beneficial effect on innovation incentives.

Several recent papers have used these and related ideas to study the efficiency effects of an independent invention defense in formal models. The starting point for welfare analysis is the observation that the defense has no effect in the absence of independent invention, and it has desirable ex post efficiency effects in the event that multiple invention does

occur because it establishes a duopoly rather than a monopoly in the use of the patented technology.[38] By immunizing Party B from charges of patent infringement, the use of the patented technology is promoted, and the deadweight loss associated with the patent is reduced. An additional ex post benefit arises in cases where Parties A and B would sign a license without the defense: the defense reduces or eliminate the transactions costs of licensing between Parties A and B.

The larger question is how the defense affects innovation incentives. The effects of the defense on innovation are captured in different ways in different formal models, but the idea that the defense promotes innovation and long-run economic efficiency is robust. In cases where the two inventors are not actual or potential competitors, the defense actually *increases* the expected rewards to innovation. If neither party can predict in advance whether it will be the first or second inventor, the shifting of rents from the first inventor to the second inventor that the defense causes will have no effect at all on their innovation incentives. Because the defense generates some ex post benefits by reducing transaction costs and by eliminating distortions associated with running royalties, it increases both parties' expected rewards.

The analysis is more complex in situations where Parties A and B are actual or potential competitors because the defense causes additional ex post competition in such cases. Nonetheless, the defense robustly improves ex ante incentives, better aligning private reward and social contribution. The most straightforward case arises if the parties' expected rewards without the defense are greater than their expected social contributions. We observed in section II.A that this is the norm when multiple firms race to obtain a patent. In this case, the independent invention defense has favorable ex ante as well as ex post effects. Maurer and Scotchmer (2002) make this point using a static model with free entry in which each firm, by paying a fixed amount, can discover the invention with certainty. In their model, all R&D expenditures by multiple firms are entirely duplicative, and the market equilibrium involves excessive entry by rent-seeking firms. The independent invention defense helps correct the common pool problem identified in section II.A.

The attractiveness of the independent invention defense is not, however, confined to situations in which the rewards to patent holders under the conventional patent system exceed their social contributions. With sufficiently large spillovers that are specific to individual innovators, a patentee's reward can, in theory, fall short of its contribution, even with

subsequent independent invention. Large, inventor-specific spillovers may not be common. However, to address this possibility, Shapiro (2006b) develops a model in which the appropriability ratio can be low.

In the model studied by Shapiro (2006b), it is efficient for multiple parties to pursue the invention in question due to uncertainty, diminishing returns to investment by each party, and the benefits associated with a diversity of approaches taken by multiple parties. He studies optimal patent policy where patent length and the availability of an independent invention defense are patent system design variables. In his model, the independent invention defense enhances long-run economic efficiency, even in the presence of large spillovers that are specific to Party A or Party B, regardless of the elasticity of supply of inventions. He shows that an independent invention defense is optimal under the mild condition that the ratio of deadweight loss to profits is higher under monopoly than under duopoly. So long as this condition is met, any given reward to successful innovation can be provided more efficiently using a longer patent lifetime combined with the independent invention defense, rather than a shorter patent lifetime without the defense.[39] By an extension of his logic, adding an independent invention defense to the current system enhances efficiency, holding fixed the patent lifetime, if the patent lifetime has been set optimally.[40] Shapiro also shows that the defense has superior properties in terms of the social portfolio of innovative projects pursued by private parties because the defense better aligns the relative private and social rewards in the event of single versus multiple invention.

In a model with free entry in which the firms do not choose their R&D investment levels, La Manna, MacLeod, and de Meza (1989) establish conditions under which a "permissive" patent regime, under which all independent inventors have the right to practice the invention so long as they invent within a specified time period after the first inventor, is optimal. Henry (2007) develops a model of "runner-up patents," in which a second inventor can freely use the patented technology if the time lag between invention by the first party and invention by the second party does not exceed some specified length. In his model, each party chooses its R&D investment level. He provides conditions under which such runner-up patents are desirable and argues that the maximum time lag during which runner-up patents would be issued should become a new patent policy tool.

As noted by Shapiro (2006b), one tricky issue surrounding the defense is that it could cause some inventors to adopt trade secret rather than

patent protection because an inventor who keeps its invention secret could subsequently invoke the defense if a patent is later issued to another party for the same invention. From a social perspective, this does not seem to present a problem in cases where the technology is, in fact, subsequently patented: in those cases, society still receives the benefit of the patent disclosure, along with the benefits of some competition in the use of the technology. Some social costs could arise in cases where the technology remains a secret for an extended period of time; those costs would need to be balanced against the various benefits of the defense discussed in the preceding.[41]

Establishing an independent invention defense would likely have salutary effects on the patenting decisions by firms, especially in areas where patent quality is most problematic. Under the current system, a party that receives a patent covering a technology that is independently developed by others can receive a very substantial reward, especially if it can use the patent opportunistically. The system, therefore, provides rewards to parties who are most aggressive in seeking patents on inventions that might be obvious or in the public domain. The defense would systematically reduce the rewards associated with such patents, while having no effect on inventors who uniquely make pioneering discoveries. The defense thus might help counteract the vicious cycle by which patents become easier to get, causing an increase in the number of patent applications, putting more pressure on the PTO, making patents even easier to get.

One objection that might be raised to an independent invention defense is that it would create some uncertainty for the patent holder about its ability to exclude rivals as it would not know whether certain rivals are in a position to assert the defense. Such uncertainty might well be present, but it is not a basis for rejecting the defense. Surely the norm in business is for parties making investments to face uncertainty about the competition they are likely to encounter in the future. Economic efficiency is promoted by aligning patentee rewards and contributions; reducing the risk faced by patent holders is not desirable per se and is not an independent goal.

Going beyond these models, one of the great benefits of the independent invention defense is that it would, in one fell swoop, do away with many of the problems with a patent holdup. A firm that independently develops the patented technology and incorporates the technology into its products would presumably qualify for the defense.[42] The independent invention defense also would go a long way to solving problems

associated with continuation applications: if a patent holder adjusts its claims during the patent prosecution process to capture products already on the market, the firms selling those products may well be able to invoke the defense successfully.[43]

B. Some Practical Considerations

The idea of an independent invention defense is not new. To begin with, such a defense already exists in other areas of U.S. intellectual property law, namely in copyright infringement cases and in trade-secret misappropriation cases.[44] Furthermore, a very limited version of the independent invention defense, called the *Early Inventor Defense*, already exists in U.S. patent law. To see how this defense works, consider an alternative fact pattern. Party A is again the patent holder. But now suppose that Party B actually invented the patented technology first. How could Party A obtain a patent with this fact pattern? Suppose that Party B did not consider this invention sufficiently novel and nonobvious to be patentable. Instead, Party B kept its invention secret, and Party A subsequently filed for a patent. According to U.S. patent law, with this fact pattern, Party A can obtain a patent.[45] Typically, Party A can then prevent Party B from practicing the invention.[46] However, in 1999, an exception to this rule was created, evidently in response to concerns that patents on business methods would be asserted opportunistically against prior users of those methods. For business method patents only, Party B can invoke the Earlier Inventor Defense if Party B "actually reduced the subject matter to practice at least 1 year before the effective filing date of such patent, and commercially used the subject matter before the effective filing date of the patent" (35 USC 273[b][1]). Congress is currently considering legislation that would greatly expand the Earlier Inventor Defense by applying it to all patents and requiring only that Party B "commercially used, or made substantial preparations for commercial use of, the subject matter before the effective filing date of the claimed invention."

Even in this expanded form, the Earlier Inventor Defense is not nearly as broad as a full-fledged independent invention defense as it can only be invoked by first inventors, not independent second inventors. Rather, the Earlier Inventor Defense is a form of "prior user rights," under which a party using an invention *before* another party applies for a patent on that invention is granted the right to continue using that invention after the patent issues. From 1836 to 1952, U.S. patent law provided for such

prior user rights, and such rights are the norm in other countries.[47] The independent invention defense would not only protect prior users; it would go further and protect independent second inventors as well.[48]

While the basic idea behind the independent invention defense is clear enough, actually introducing such a defense would inevitably raise a host of practical issues. A comprehensive practical and legal discussion of these issues is beyond the scope of this paper. Fortunately, many of the legal issues have been addressed by Leibowitz (2002), Armond (2003), Lemley (2007), and, in greatest depth, by Vermont (2006), who strongly advocates a "reinvention" defense. Here, we confine our attention to two of the most significant practical issues where the policy analysis is greatly informed by economic reasoning.

First and foremost, the defense should only apply if the second inventor *independently* achieves the invention that it practices. In principle, independence is clear enough, but in practice, it may be very difficult to ascertain. To fix ideas, consider the following timing patterns of invention by Party A, the patent holder, and Party B, who is invoking the defense:

1. Party B achieved the invention before Party A but concealed its invention.[49]

2. Party B achieved the invention after Party A, but before the public disclosure of Party A's invention.

3. Party B achieved the invention after Party A and after the public disclosure of Party A's invention.

The public disclosure envisioned in (2) and (3) must be sufficiently detailed to teach a person of ordinary skill in the art how to make and use the invention. Public disclosure could come in the form of affirmative publication by Party A, the disclosure of Party A's patent application by the PTO, or the issuance of the patent itself. A full-fledged independent invention defense would apply in all three of these timing patterns.

Establishing the defense under timing pattern (1) appears straightforward. Under timing pattern (1), there would presumably be no doubt that Party B is indeed an independent inventor. This is the timing pattern under which the Early Inventor Defense applies. Timing pattern (1) could also be handled by granting "prior user rights" to Party B, the first inventor. As noted in the preceding, many European countries already grant prior user rights to parties who practice an invention secretly, prior to the date at which another party applies for a patent on that in-

vention. This system appears to work well in Europe, and there does not appear to be any reason to fear that granting such rights would cause significant practical problems in the context of the U.S. patent system.

Extending the defense to timing pattern (2) also appears to be practical. This would, in effect, grant prior user rights to parties who achieve an invention prior to the date at which another party discloses that invention (but not prior to the other party's invention). In principle, this appears to be a relatively minor step, but a helpful one in terms of aligning rewards with contributions and avoiding patent holdup. Under this timing pattern, Party B might be afforded a presumption that it is an independent inventor. However, Party A could rebut this presumption by presenting evidence showing that Party B received proprietary information about Party A's invention prior to its public disclosure. Actual notice of the invention provided by Party A to Party B prior to Party B's invention would also rebut the presumption. This approach would have the advantage of encouraging Party A to disclose its invention at an early date (either publicly or on a selective basis) in order to limit the ability of others to invoke the defense. Restricting the defense to timing patterns (1) and (2) would also give Party B an incentive to remain well informed about inventions that have been publicly disclosed and to focus its efforts on novel inventions, rather than devoting resources in pursuit of inventions that have already been achieved.[50]

Establishing a full-fledged independent invention defense, including timing pattern (3), would raise some thorny additional issues that might prove insurmountable. With this timing pattern, it would be especially important to place the burden of proof on the party asserting the defense. For this very reason, there may be limited scope for the defense to come into play as it may be difficult for Party B to establish that its invention truly was independent of Party A. Party B might produce lab notebooks indicating that it achieved the invention independently. But Party B might have difficulty proving that it did not learn something from the public disclosure of Party A's invention. Sanctions for fraud would surely limit Party B from blatantly lying about the receipt of such information, but there would undoubtedly be cases that were not clear-cut. Plus, extending the defense to this timing pattern would create a disincentive to reading patents.

A secondary practical issue involves the organizational scope of the defense. If Party B were allowed to extend the defense to third parties by contract, Party B would effectively have the right to issue licenses to use the patented technology. A defense with this feature could have eco-

nomic effects closer to placing the patented technology in the public domain than would a defense without licensing rights. Even if the defense cannot be licensed in this way, the question still arises as to whether the party invoking the defense can extend its rights to other parties by assignment or through acquisition. Current patent law addresses this issue. The Earlier Inventor Defense is only a "personal defense." The defense can only be assigned as part of an entire enterprise, and in that case, a limitation applies to the sites at which the defense may be asserted.

IV. Enhanced Patent Reexamination

We now explore the economic effects of patent reexaminations using the reward/contribution framework developed in the preceding. Reexaminations are potentially an important method to weed out improperly granted patents before they cause economic harm.

The problem of patent quality is not a new one, but it has taken on increased importance with the explosion of patenting, especially in the information-technology sector and with widespread concerns over the ability of the PTO to identify relevant prior art, given that nearly 450,000 patent applications were filed at the PTO during fiscal year 2006.[51] The normal patent examination procedure is not an effective way to elicit information on prior art. Applicants are not required to conduct a search for prior art, and third parties, who often possess a great deal of information about prior art along with the incentive to inform the PTO about the prior art, have little or no role to play. While improvements in the initial examination process are no doubt possible, many patents are of no commercial significance, and a significant fraction are not even renewed.[52] Therefore, subjecting all patent applications to a very thorough review is not sensible or practical, as emphasized by Lemley (2001).

In principle, a post-grant review procedure at the PTO, applied selectively to questionable patents with commercial significance, can correct for these deficiencies in the initial examination process. Under current U.S. patent law, patent reexaminations are meant to serve this purpose. The vast majority of reexaminations are *ex parte* reexaminations; these can be initiated by the patent holder, by the PTO, or by third parties who then play no ongoing role. Almost half of all *ex parte* reexaminations are initiated by the patent holder, usually to strengthen the patent in the face of new information about prior art.[53] A very small number of reexaminations are *inter partes* reexaminations, in which the challenger can participate actively.[54]

The patent reexamination system in the United States does not appear to be operating very effectively.[55] There is a growing consensus that improvements are needed in the system by which issued patents can be challenged at the PTO. Recommendation number 1 by the FTC (2003) states: "As the PTO Recommends, Enact Legislation to Create a New Administrative Procedure to Allow Post-Grant Review of and Opposition to Patents." The National Academies of Science (NAS; 2004, 95–103) recommends an "Open Review Procedure" to replace existing *inter partes* reexamination and *ex parte* reexaminations initiated by third parties. All of the patent reform proposals currently being considered in Congress include enhanced post-grant opposition procedures. These proposals raise a number of practical issues that are beyond the scope of this paper. Here, we show how an enhanced reexamination procedure would help align patentee rewards and contributions.

A. Key Attributes of Patent Reexamination Procedure

For our purposes, there are two key attributes of any enhanced reexamination procedure.

First, the procedure should allows third parties with significant relevant information to initiate reexamination at an early date, before the patent is licensed and before users make investments specific to the patented technology. In this respect, reexamination is very different from patent litigation, which takes much longer, and where the patent holder tends to control the timing.[56]

Second, the procedure must elicit significant information about patent validity or scope, causing a meaningful fraction of the patents that are reexamined to be invalidated or narrowed. This is the primary function of the reexamination procedure. However, as emphasized in Farrell and Shapiro (2008), these cannot be the only effects of the reexamination procedure. The laws of probability imply that the patents surviving reexamination must emerge *stronger:* what does not kill a patent makes it stronger. For example, suppose that a patent has strength of 30 percent, that is, there is a 30 percent chance that it will be held valid if litigated. If subjected to reexamination, suppose that there is a 50 percent chance the patent will be found invalid and a 50 percent chance the patent will survive the reexamination. The basic laws of probability then imply that the patent will have a strength of 60 percent if it survives the reexamination.[57] This observation is very important in the following: if some

patents are weakened or invalidated by reexamination, others must be strengthened.

B. Aligning Rewards and Contributions

To understand the economic effects of enhanced reexaminations, we begin by asking how such reexaminations differ from patent litigation, the established procedure for testing patents. Two differences are clear, and we will not dwell on them. First, reexamination is far less expensive than litigation. Second, the strong presumption of validity that applies in patent litigation need not apply in a reexamination. To see how reexaminations affect the alignment between patentee rewards and contributions, we focus on the difference in timing: reexaminations can occur shortly after the patent is issued and at the instigation of parties other than patent holder.

In section II.B, we showed how excessive rewards to probabilistic patents can result from the public good nature of patent challenges. Farrell and Shapiro (2008) show that an expanded reexamination procedure could substantially reduce these excessive rewards for patents that are licensed to multiple downstream rivals. We explained in section II.B why any one downstream firm has a limited incentive to challenge the patent once it has been licensed to a number of the firm's rivals: if the patent is upheld, the challenger is at a competitive disadvantage, and if the patent is invalidated, the challenger gains no competitive advantage. This "relativity" logic is very different if a downstream firm can initiate a reexamination *before* the patent has been licensed: whether the patent is upheld or invalidated, the challenger will be on an equal footing with its rivals. While there will remain a public good aspect to patent challenges, each individual downstream firm has a much stronger incentive to challenge a patent before it has been licensed to its rivals. Allowing technology users to initiate reexaminations limits the ability of patent holders to strategically control the timing of licensing and patent litigation.

Because the costs associated with improperly issued patents are also borne by final consumers, there are benefits from allowing entities that represent their interests (but may lack standing to initiate a court challenge of the patent) to initiate a reexamination, at least if they can present significant and credible new evidence to the PTO. There already exist some such entities, and their role might expand if the rules were changed to encourage more patent reexaminations.[58]

An enhanced reexamination procedure would also help align rewards and contributions by reducing the ability of owners of probabilistic patents to engage in patent holdup. In section II.C, we showed how excessive rewards to probabilistic patents can result from patent holdup, even if the alleged infringers are fully aware of the patent when they make their product design decisions. As shown in Shapiro (2006a), a downstream user facing an infringement claim from the owner of a weak patent may have no way to prevent patent holdup unless it can effectively test the patent's validity (or scope) at an early date, before it makes technology-specific investments. Patent litigation typically does not afford the user this option. If the downstream firm's only recourse to taking a license is to either avoid infringing or to produce an infringing product and then bear the risk of patent litigation, the patent holder will predictably capture some holdup rents. A strong reexamination procedure would give users a way of testing patent strength at an early date, bringing the patent holder's return closer in line with its contribution.

V. Additional Reforms to Align Reward and Contribution

We now briefly consider three other reforms to the patent system to illustrate the utility of viewing patent system design through the rewards/contribution lens.

A. Limiting Injunctions in Selected Infringement Cases

The Supreme Court recently held in *eBay, Inc. v. MercExchange* that district courts have the power to deny permanent injunctions in appropriate patent infringement cases.[59] This ruling was a departure from the approach previously taken by the Federal Circuit, under which permanent injunctions were virtually automatic following a finding of infringement.[60] We showed in section II.C that the threat of obtaining a permanent injunction can allow owners of probabilistic patents to obtain rewards that exceed their (expected) contributions. Excessive rewards are especially likely in cases where the patent covers a minor feature in a complex product with a large price/cost margin. Limiting the use of injunctions can correct for these excessive rewards. Lemley and Shapiro (2007) offer a procedure to determine whether an injunction should be granted, based on the costs of redesigning the product to avoid infringing and the time lag involved in such redesign. In cases where the patent holder does not compete against the infringing firm, and where the in-

fringing firm developed the technology independently of the patent holder, they propose that injunctions be denied if the redesign costs are large relative to the value of the patented technology, and delayed if these costs are modest but involve a significant time lag. Following the logic from Shapiro (2006a), these proposals are designed to prevent patent holders from capturing excessive returns based on holdup, while protecting the ability of patent holders to earn rewards commensurate with their inventive contributions, regardless of whether they exploit their intellectual property internally or through licensing.

The analysis on complementary innovations in section II.D also is directly relevant for evaluating the rules regarding injunctions. Even if the injunction threat does not provide a reward to the patent holder that exceeds its contribution, it can disadvantage other innovators whose investments are subject to holdup, thus reducing efficiency and retarding innovation.

B. Reasonable Royalties in Components Cases

In patent infringement cases, courts often must determine the "reasonable royalties" damages that the infringing party must pay to the patent holder. As discussed in detail by Lemley and Shapiro (2007), there is a danger that court-determined reasonable royalty rates will be excessive in cases where the patent covers only a minor feature of a complex and valuable product. They present some evidence from actual damages awards that such errors are being made. The problem is sharply illustrated by the recent case involving MP3 patents and Microsoft's Media Player.[61] In that case, a jury awarded the patent holder Alcatel-Lucent $1.52 billion in damages; this award was later reversed on other grounds. Alcatel-Lucent had argued that damages should be based on 0.5 percent of the total value of computers sold with Microsoft Windows (which includes the Media Player). While 0.5 percent might seem like a small percentage royalty, it was applied to a very large base of sales, despite the fact that Windows computers have literally thousands of components at the level of the patented MP3 software elements. Due to the enormous specific investments made in the MP3 standard, there is considerable scope for patent holdup and excessive royalties in this type of situation.

In principle, the courts could correct for this problem by making it clear that royalties should be based on the underlying value of the patented feature, in comparison with the best noninfringing ex ante alternative, and not based on the entire value of the infringing product. In

fact, the Supreme Court made just such a statement over 150 years ago.[62] The patent reform bill that passed the House of Representatives on September 11, 2007 (HR 1908, Section 5(b)) includes the following language to address this problem:

(2) RELATIONSHIP OF DAMAGES TO CONTRIBUTIONS OVER PRIOR ART- Upon a showing to the satisfaction of the court that a reasonable royalty should be based on a portion of the value of the infringing product or process, the court shall conduct an analysis to ensure that a reasonable royalty . . . is applied only to that economic value properly attributable to the patent's specific contribution over the prior art. The court shall exclude from the analysis the economic value properly attributable to the prior art, and other features or improvements, whether or not themselves patented, that contribute economic value to the infringing product or process.

(3) ENTIRE MARKET VALUE- Upon a showing to the satisfaction of the court that the patent's specific contribution over the prior art is the predominant basis for market demand for an infringing product or process, damages may be based upon the entire market value of the products or processes involved that satisfy that demand.

Another recent patent reform bill includes the following language to address this problem:

In determining a reasonable royalty consideration shall be given to (A) the economic value that should be attributed to the novel and non-obvious feature or features of the invention, as distinguished from the economic value attributable to other features, improvements added by the infringer, and the business risks the infringer undertook in commercialization; [and other factors]. (*Hatch-Leahy*, S3818)

Because patent royalties are negotiated in the shadow of litigation, insuring that "reasonable royalty" awards are not excessive would have a ripple effect on the far larger number of cases in which royalties are negotiated to avert or settle litigation.

C. The Doctrine of Willful Infringement

A party found to have willfully infringed a patent can be subject to treble damages. While this doctrine may be useful to deter infringing parties from intentionally copying patented inventions, if interpreted too broadly it can provide excess rewards to patent holders. Lemley and Tangri (2003) criticize the doctrine of willful infringement and propose reforms to it, including evaluating willfulness based on the situation when

the alleged infringer first adopted the patented technology, rather than treating subsequent continued use of the patented technology, after the infringer received notice from the patent holder, as willful infringement.[63] They also note that, to the extent that willfulness is based on whether and when the downstream firm actually read the patent, it deters users from reading patents and disrupts the patent disclosure function.

Reform of the willfulness doctrine can be informed by the reward/contribution approach. To see how, consider a downstream manufacturer who has already designed its product and is then accused of patent infringement. Even if the manufacturer did not independently invent the patented technology, it may well have a reasonable, good faith belief that the patent is either invalid or not infringed. The manufacturer may well hold these beliefs even after receiving a threat letter from the patent holder. For example, the manufacturer might believe that the patent is valid and infringed with a 60 percent probability; the patent holder might even hold the same beliefs. Treble damages are likely to undermine rather than promote economic efficiency in this circumstance. Exposing the downstream firm to treble damages if it continues infringing will tend to exacerbate the patent holdup problem inherent in this fact pattern. One recent patent reform bill includes the following language to reform the doctrine of willful infringement:

> A court shall not find that an infringer has willfully infringed a patent . . . for any period of time during which the infringer had an informed good faith belief that the patent was invalid or unenforceable, or would not be infringed by the conduct later shown to constitute infringement of the patent. (*Hatch-Leahy*, S3818)

As with "reasonable royalties," reforming the willfulness doctrine would have a ripple effect on the far larger number of cases in which licenses are signed to avert or settle patent litigation. The Federal Circuit Court of Appeals recently took a significant step in this direction, establishing a new "objective recklessness" standard for willful infringement in *In Re Seagate Technology, LLC*, 497 F.3d 1360, August 20, 2007.

VI. Conclusions

The complexity of the issues involved in reforming the patent system can make patent reform appear to be an intractable problem. On one side are those who stress the enormous social benefits flowing from innovation and the attendant risk that reducing the rewards to patent holders will unwisely discourage socially beneficial innovation. This side stresses inno-

vation spillovers and the importance of well-defined property rights for commercialization. On the other side are those who see a broken system that often rewards patent holders who have invented little, if anything, at the expense of others who are themselves making the investments and taking the risks necessary for innovation. This side stresses defects in the system by which patents are grants and the leverage afforded to patent holders, especially in the information technology industry.

This paper argues that considerable clarity regarding patent reform can be achieved by viewing proposed reforms through the reward/contribution lens: reforms that better align rewards and contributions, and especially those that prevent patent holders from earning rewards in excess of their social contributions, raise economic efficiency and promote innovation, regardless of the (unobservable) elasticity of supply of innovations. Introducing an independent invention defense in patent infringement cases would predictably reduce excess rewards. An independent invention defense could encourage greater diversity in innovation, which could generate considerable benefits. Strengthening the procedures by which patents are reexamined after they are issued would also help align rewards and incentives, especially for questionable patents and patents licensed to multiple competing firms. Several other proposed reforms relating to patent infringement litigation also look promising using the reward/contribution framework: limiting the use of injunctions, clarifying the way in which damages based on "reasonable royalties" are defined, and narrowing the doctrine of willful infringement. Recent empirical work suggests that all of these reforms would affect important, real-world situations that frequently arise. The economic reasoning supplied here suggests that these reforms would enhance economic efficiency and promote innovation.

Appendix

Two Simple Models in which Full Appropriation Is Optimal

Suppose that a single firm chooses its R&D budget aimed at a specific, well-defined invention. This is the only firm capable of pursing this invention. The greater is the firm's investment, the higher is the probability of achieving the invention. Formally, the firm can achieve a probability of success p by spending an amount $C(p)$, with $C'(p) > 0$ and $C''(p) > 0$ for all $0 < p < 1$.

As an illustration, consider the situation in which the firm can decide how many separate projects to pursue toward the invention. If each

project has a probability q of succeeding, and if all of the projects are statistically independent, then the probability of succeeding if N projects are pursued is $p = 1 - (1 - q)^N$, so $N = \log(1 - p)/\log(1 - q)$. If each project costs k to pursue, the cost function is given by $C(p) = kN = k \log(1 - p)/\log(1 - q)$.

Let the social contribution of the invention be V. The expected social net benefits are given by $pV - C(p)$. The social optimum is achieved by picking a success probability p^*, which maximizes this expression; p^* is characterized by $C'(p^*) = V$.

Suppose that the firm captures profits π if it achieves the invention. If the firm maximizes its expected profits, it will choose p to maximize $p\pi - C(p)$, so it will choose the "monopoly" probability p_M that satisfies $C'(p_M) = \pi$. Clearly, the firm will spend the socially optimal amount pursuing this invention if and only if $\pi = V$, that is, with full appropriation. If $\pi < V$, the invention causes positive externalities, that is, spillovers, and $p_M < p^*$; conversely, if $\pi > V$, the invention causes negative externalities, and $p_M > p^*$. This simple model, therefore, suggests that it is desirable to reward the firm with the full social contribution associated with the invention. This idea is rooted in the core economic concept of externalities.

A very similar result can be obtained in another simple model that focuses on the timing rather than the probability of invention. Suppose that the firm can achieve the invention at time T with an expenditure of $E(T)$, where $E(T)$ and all costs and benefits in this model are measured in present discounted value as of time zero. Suppose that the social contribution of the invention is $V(T)$ if it is achieved at time T. The social net benefits are $V(T) - E(T)$, which we assume are concave in T. The optimal date of invention T^* is characterized by $V'(T^*) = E'(T^*)$.

Now suppose that a single firm is choosing how much to spend to achieve this invention, and the firm's profit if it achieves the invention at date T is $\pi(T) = \alpha V(T)$. This firm will pick T to maximize $\pi(T) - E(T)$, leading to invention date T_M, defined by $\alpha V'(T_M) = E'(T_M)$. With full appropriability, $\alpha = 1$, the firm's choice is socially optimal, $T_M = T^*$. With partial appropriability, $\alpha < 1$, the firm spends less than the social optimum on the invention, and discovery is delayed: $T_M > T^*$. Likewise, with more than 100 percent appropriability, $\alpha > 1$, the firm spends more than the social optimum on the invention, $T_M < T^*$.

Social Contribution Associated with a Probabilistic Patent

Consider a technology that lowers the cost of producing a given product from c to $c - v$. Let the demand for this product be given by $D(p)$,

where p is the downstream price. The consumer surplus function corresponding to the demand function $D(p)$ is $S(p) \equiv \int_p^\infty D(q)dq$. For simplicity, assume that the downstream industry is perfectly competitive, so price equals cost.

If the patented technology is not available, the downstream price is $p = c$, output is $D(c)$, and total welfare is given by $S(c)$. By similar reasoning, if the patented technology is freely available, $p = c - v$, output is $D(c - v)$, and total welfare is given by $S(c - v)$.

If the patent holder licenses its patent to the downstream industry as a per-unit royalty of $r \leq v$, the downstream price will be $p = c - v + r$, and output will be $D(c - v + r)$. The total welfare in this case is the sum of consumer surplus $S(c - v + r)$, and the patent holder's royalty revenues, $rD(c - v + r)$, so total welfare is given by $W(r) \equiv S(c - v + r) + rD(c - v + r)$.

Let the patent strength be θ, where $0 < \theta < 1$. If the patent is licensed at a per-unit royalty rate r, then the patent holder's contribution is equal to the difference between $W(r)$ and the expected welfare that would have resulted in the absence of the patent holder, which is equal to $\theta S(c) + (1 - \theta)S(c - v)$. Why? With probability θ, the patent holder truly invented the patented technology, so without the patent holder, price would be $p = c$, and welfare would be $S(c)$. Likewise, with probability $1 - \theta$, the patented technology was anticipated in the prior art (or obvious), so without the patent holder, price would be $p = c - v$, and welfare would be $S(c - v)$.

Therefore, the contribution of the patent holder equals $W(r) - [\theta S(c) + (1 - \theta)S(c - v)]$, which is $S(c - v + r) + rD(c - v + r) - [\theta S(c) + (1 - \theta)S(c - v)]$. Because the profits of the patent holder are $rD(c - v + r)$, the gap between contribution and profits is $S(c - v + r) - [\theta S(c) + (1 - \theta)S(c - v)]$.

Following Farrell and Shapiro (2008), theorem 3, we now show that if $r > \theta v$, this gap is negative, so the patent holder's profits exceed its contribution. This requires showing that consumers are worse off paying the price $c - v + r$ for sure rather than paying the price c with probability θ and the price $c - v$ with probability $1 - \theta$. The expected price in the absence of the patent holder is $\theta c + (1 - \theta)(c - v)$, which equals $(c - v) + \theta v$. The price with the patent holder is $(c - v) + r$. If $r > \theta v$, then the patent holder's presence causes the expected price to rise. Furthermore, the patent holder's presence eliminates price risk: rather than face a gamble between a price of c and a price of $c - v$, with probabilities θ and $1 - \theta$, respectively, consumers pay the price $c - v + r$ for sure. Standard consumer theory tells us that consumers are risk loving in price: $S'(p) = -D(p)$

and $S''(p) = D'(p) > 0$, so the elimination of price risk makes consumers worse off. Therefore, the presence of the patent holder makes consumers worse off: the patent holder raises the expected price and reduces (attractive) price risk.

We can quantify the negative effect of the patent holder on consumers. For small or moderate values of v, $S(c - v + r) \approx S(c) - (v - r)S'(c)$ and $S(c - v) \approx S(c) - vS'(c)$, so the gap between contribution and profits is (approximately) $S'(c)[r - \theta v]$. Since $S'(c) = -D(c)$, this is equal to $-D(c)[r - \theta v]$. Therefore, the gap between profits and contribution is (approximately) given by $D(c)[r - \theta v]$. For small values of v, the patent holder's excess payoff is equal to the product of two terms: (1) the difference between the per-unit royalty and the expected per-unit costs savings attributable to the patent holder, and (2) the number of units sold.

The gap between profits and contribution can usefully be expressed in terms of downstream industry revenues, which equal $(c - v + r)D(c - v + r)$. Again for small values of v, these are approximately equal to $cD(c)$. The ratio of the profit/contribution gap to industry revenues is thus (approximately) given by $D(c)[r - \theta v]/cD(c)$ or $(r - \theta v)/c$. Because $p = c - v + r \approx c$, this ratio is roughly equal to $(r - \theta v)/p$. If the downstream price is \$100, the patent strength is 30 percent, the value of the patented technology is \$10, and the royalty rate is \$5, then the excess reward is equal to $(5 - 0.3 \cdot 10)/100$, or 2 percent of industry revenues.

Complementary Innovations

Suppose that firm i can achieve a probability of invention of p_i with an expenditure of $C_i(p_i)$, and one firm's success is independent of the other. Suppose that each firm, i, fully appropriates its social contribution, V_i, if it is the only successful innovator. (This is optimal in this model and does not require any subsidies.) How should the social contribution of V_B be allocated if both innovate? Let X denote the portion of S awarded to firm 1, with $S - X$ awarded to firm 2.

Firm 1's expected profits are $p_1p_2(V_1 + X) + p_1(1 - p_2)V_1 - C_1(p_1)$, so the first-order condition for p_1 is $C_1'(p_1) = p_2(V_1 + X) + (1 - p_2)V_1$. Note that $dp_1/dX = p_2/[C_1''(p_1)]$. Likewise, firm 2's expected profits are $p_1p_2(V_2 + S - X) + p_2(1 - p_1)V_2 - C_2(p_2)$, so the first-order condition for p_2 is $C_2'(p_2) = p_1(V_2 + S - X) + (1 - p_1)V_2$, and $dp_2/dX = -p_1/[C_2''(p_2)]$.

The expected social contribution is $p_1p_2V_B + p_1(1 - p_2)V_1 + p_2(1 - p_1)V_2 - C_1(p_1) - C_2(p_2)$. Differentiating this expression with respect to X and set-

ting the resulting expression equal to zero and using the first-order conditions for p_1 and p_2, gives $(dp_1/dX)p_2(S - X) + (dp_2/dX)p_1X = 0$. Converting this into elasticity form gives

$$\left(\frac{X}{S-X}\right)^2 = \frac{\varepsilon_1}{\varepsilon_2},$$

where ε_i is the elasticity of firm i's probability of success with respect to the portion of the synergy it receives in the event that both firms succeed. The right-hand side of this expression measures the relative sensitivity of firm 1's versus firm 2's probability of success to the portion of the synergy that firm captures. The optimal division of the synergy reflects this relative sensitivity. This expression is similar to standard rules for Ramsey pricing.

With symmetry and the appropriate concavity conditions, it is optimal for each of the two firms to appropriate half of the synergy in the event they are both successful. With symmetry and concavity among N complementary innovations, the optimal appropriation share under a system without subsidies will simply be $1/N$. A system that rewarded the lion's share of the synergies to one firm would be less efficient.

Accounting for the Deadweight Loss Caused by Patent Monopolies

We now return to the model in which a single party can achieve a given innovation with probability p according to the cost function $C(p)$. We now suppose that stronger patent rights not only allow the firm to capture greater profits from the invention but also generate greater deadweight loss. One simple, reduced-form way to model such situations is by introducing the function $V(\pi)$ to represent the social contribution of the invention if the patent system allows the firm to capture profits π from the invention. Here $V(0)$ represents the social contribution from the invention if it is not subject to patent protection. Stronger patent protection allows the patent holder to capture greater profits but also leads to greater deadweight loss, so $V'(\pi) < 0$.

The firm chooses p to maximize $p\pi - C(p)$; we denote the firm's optimal choice by p_M, which is determined by the first-order condition $C'(p_M) = \pi$. Note that

$$\frac{dp_M}{d\pi} = \frac{1}{C''(p_M)} > 0.$$

We now treat π as a reduced-form policy variable representing the strength of patent rights. The social net benefits as a function of π are $p_M V(\pi) - C(p_M)$. So long as this expression is concave in π, maximizing these social net benefits with respect to π gives the first-order condition $p_M V'(\pi) + V(\pi)(dp_M/d\pi) - C'(p_M)(dp_M/d\pi) = 0$. Using $C'(p_M) = \pi$, this becomes $p_M V'(\pi) + [V(\pi) - \pi](dp_M/d\pi) = 0$. Because $V'(\pi) < 0$ and $dp_M/d\pi > 0$, this implies that $\pi < V(\pi)$, that is, partial appropriation is socially optimal. Manipulating the expression for the optimal π gives

$$\frac{V(\pi) - \pi}{V(\pi)} = \frac{(-dV/d\pi)(\pi/V)}{(dp_M/d\pi)(\pi/p_M)}.$$

This equation tells us that the fraction of social benefits *not* appropriated by the firm is equal to the ratio of two elasticities: the elasticity of social contribution with respect to the firm's profits (numerator) and the elasticity of the probability of invention with respect to the firm's profits (denominator).

Endnotes

I owe a special debt to Joe Farrell, who has been invaluable in helping me better understand many of the issues addressed here and who generously commented on previous drafts of this paper. I also thank Richard Gilbert, Bronwyn Hall, Josh Lerner, and Scott Stern for their valuable comments on an earlier draft.

Carl Shapiro is the Transamerica Professor of Business Strategy, Haas School of Business, and Professor of Economics, University of California at Berkeley. My papers cited here are available on my Web site, http://faculty.haas.berkeley.edu/shapiro.

1. See Federal Trade Commission (2003), National Academies of Science (2004), and Jaffe and Lerner (2004).

2. The National Academies of Science (2004) emphasized the need to reinvigorate the nonobviousness standard. In April 2007, in the *KSR International v. Teleflex Inc.* case, the Supreme Court issued a major ruling that defined what is "obvious," and thus not patentable, more broadly than had the Federal Circuit Court of Appeals.

3. The Federal Trade Commission (2003) emphasized the problem of "questionable" patents, and the National Academies of Science (2004) made suggestions for improving patent quality. Jaffe and Lerner (2004) also stress the problem of patent quality. Lemley and Shapiro (2005) summarize some of the evidence regarding patent quality.

4. In a key decision, *Kingsdown Medical Consultants v. Hollister,* 863 F.2d 867 (1988), the Federal Circuit stated:

It should be made clear at the outset of the present discussion that there is nothing improper, illegal or inequitable in filing a patent application for the purpose of obtaining a right to exclude a known competitor's product from the market; nor is it in any manner

improper to amend or insert claims intended to cover a competitor's product the applicant's attorney has learned about during the prosecution of a patent application.

Lemley and Moore (2004) discuss abuses involving continuation applications and report that over half of all litigated patents result from continuation applications. Graham and Mowery (2004) and Hegde, Mowery, and Graham (2007) provide extensive empirical evidence regarding the use of continuation applications.

5. See, for example, Shapiro (2006a) and Lemley and Shapiro (2007).

6. Examples include the Federal Trade Commission's (FTC) recent case against Rambus and a complex of cases involving Qualcomm. See Farrell et al. (2007) for an extensive discussion of patent holdup in the standard-setting context.

7. Problems with the willfulness doctrine are discussed in Lemley and Tangri (2003).

8. See Lemley and Shapiro (2007).

9. As Joseph Stiglitz, then Chairman of the Council of Economic Advisors, testified at the FTC Hearings on Global and Innovation-Based Competition on October 12, 1995:

We often talk about how important patents are to promote innovation, because without patents, people don't appropriate the returns to their innovation activity, and I certainly very strongly subscribe to that. The key importance of intellectual property rights is part of the mechanism that the market economy has to stimulate motivation. It was so important that it was included in the Constitution, so it gives you a sense of how important that is. On the other hand, some people jump from that to the conclusion that the broader the patents rights are, the better it is for innovation, and that isn't always correct, because we have an innovation system in which one innovation builds on another. If you get monopoly rights down at the bottom, you may stifle competition that uses those patents later on, and so the breadth of utilization, maybe I should say, in a broader sense, the breadth and utilization of patent rights can be used not only to stifle competition, but also have adverse effects in the long run on innovation. We have to strike a balance.

10. Patent breadth, like patent length, is a relatively crude instrument for adjusting the rewards to patent holders. However, Gilbert and Shapiro (1990) and Klemperer (1990) studied the trade-off between length and breadth in rewarding patent holders, which allows one to consider the optimal "shape" of patents, for any given level of total reward, again finessing the question of the elasticity of supply of inventions. Tandon (1982) studied a similar trade-off in providing rewards to innovators, but his analysis was cast in terms of compulsory licensing of patents rather than the scope of patent protection. Gallini (1992) considers length versus breadth in the presence of imitation.

11. There is a sizeable literature attempting to measure the technological spillovers associated with innovations. Mansfield et al. (1977) and Griliches (1979) are classic references; Hall (1996) surveys this evidence.

12. For a clear discussion of some of these issues, see Scotchmer (2004), especially chapter 4, "On the Optimal Design of Intellectual Property."

13. In general, if a patent is to be awarded to a single firm, the optimal appropriation ratio, while less than unity, depends in a complex way on the distribution of firms' costs of pursuing the invention, the likelihood of successful invention by the various firms, and the relationship between the patent holder's reward and deadweight loss. None of these variables is observable in practice.

14. Each party's incremental social contribution is nil if the two parties achieve the very same invention. In practice, different parties may make slightly different discoveries. The logic provided here suggests the virtues of a system that rewards each for its distinctive contribution, but not for any common elements.

15. Such a system would not generally be optimal as it would not provide sufficient rewards to independent *second* inventors: their social contribution would be positive (the elimination of the monopoly deadweight loss that occurs when the invention is placed in the public domain), but they would not be fully rewarded for these contributions. Shapiro (2006b) shows that optimality requires that the ratio of profits to contribution for the first inventor equal the ratio of profits to contribution for the second inventor.

16. In contrast to the discussion in section II.A, we assume here that if the patented technology is novel and nonobvious, no other entity would have discovered the patented technology, had the patent holder not done so.

17. Their analysis easily extends to the case where the patented technology adds value v to the downstream product.

18. See theorem 3 in Farrell and Shapiro (2008). The theorem relies on the mild assumption that per-unit royalties are passed through in a linear or concave manner to downstream prices.

19. Their welfare analysis is sharply different from that in Anton, Greene, and Yao (2006), who implicitly credit the patent holder with the invention at issue in their analysis of the implications of weak patent rights.

20. The U.S. Supreme Court has ruled that if one challenger to a patent prevails on patent invalidity, other users can rely on this result and, therefore, need not pay royalties, even if they had previously agreed to do so. See *Blonder-Tongue Labs, Inc. v. University of Illinois Foundation*, 402 U.S. 313, 350 (1971).

21. Farrell and Shapiro (2008) provide an extensive discussion of the issues that arise in this context, including the possibility that the downstream firms will coordinate their efforts to challenge the patent. Their analysis assumes that the patent holder has a credible threat to litigate against a single downstream firm that does not sign a license and infringes. Further work is needed to integrate the forces of relativity and litigation credibility.

22. The gap between the patent holder's profits and its social contribution for weak patents is further increased if it is feasible for the patent holder to license its patent using two-part tariffs that include negative fixed fees. In that case, the patent owner can construct a set of patent licenses that raise joint profits by supporting a hub-and-spoke cartel among the downstream firms. The patent holder makes a fixed payment to each downstream firm to induce it to sign a license rather than challenge the patent.

23. Their analysis is also relevant to assessing the benefits of improving patent quality by examining patent applications more closely *before* patents are issued. However, we apply these ideas to reexaminations because it would not be practical to greatly expand examination of all patent applications, as emphasized by Lemley (2001).

24. All of these results are obtained without any specific investments and patent holdup. As discussed in what immediately follows, patent holdup provides a distinct reason why reexaminations can generate significant social benefits.

25. See Shapiro (2006a) and Farrell et al. (2007) for simple models of patent holdup. Shapiro considers probabilistic patents; Farrell et al. consider a patent that is known to be valid.

26. If the patent holder has all of the bargaining power, the appropriation ratio without holdup would be 100 percent ($20/$20) and with holdup it would be 300 percent ($60/$20).

27. See Lemley and Shapiro (2007) and Shapiro (2006a). The problem of patent holdup can be especially severe in a standard-setting context, as discussed in Farrell et al. (2007).

28. The availability of *preliminary* injunctions also can tilt bargaining power in favor of the patent holder and lead to excessive returns, but they are rarely issued. See Armond (2003) for an extensive discussion of the independent invention defense in situations where patent holders seek preliminary injunctions.

29. To steer clear of the issues raised in the preceding section regarding independent invention of any single invention, we assume here that each of these firms is uniquely capable of achieving its respective invention.

30. This idea is a close cousin of the classic Ramsey pricing analysis, where fixed costs are most efficiently covered with markups on goods with relatively inelastic demand. These ideas are standard in optimal taxation theory.

31. Gilbert and Katz (2007) discuss how best to divide up the total payoff between firms engaging in R&D in the polar case, where no value is created until all of a fixed number of inventions are achieved.

32. See especially chapter 5, "Standing on the Shoulders of Giants: Protecting Cumulative Innovation" and the references provided there.

33. Gallini and Scotchmer (2002) emphasize that the optimal design of intellectual property in the context of cumulative innovation hinges on the effectiveness of such ex ante contracting.

34. An exception would arise if the patent holder could engage in perfect price discrimination. While this is a useful theoretical benchmark, it is not a reasonable assumption in practice, especially when one recognizes that the deadweight loss associated with a patent comes not just from higher prices, which discourage usage of the invention, but also from the reduced incentives of third parties to develop complementary innovations and improvements that build on the patented invention.

35. Congress is currently considering proposals to shift from a first-to-invent system to a first-to-file system, which is used elsewhere in the world. The presence and nature of an independent invention defense is logically separate from the rule establishing invention priority.

36. See Cohen and Ishii (2005) for a study of interference actions at the PTO. They find that interferences are highly concentrated among chemical and biomedical firms.

37. Merton (1957) is a classic reference on the frequency of multiple discovery in science. See Lamb and Easton (1984) for an in-depth treatment of multiple discoveries in the sciences.

38. The duopoly is not symmetric: only the patent holder has the right to sue others for patent infringement and the right to grant licenses to others to use the patented technology. Establishing an independent invention defense thus leads to a different system from the one in which the second inventor shares the patent and has licensing rights.

39. Ayres and Klemperer (1999) make a similar argument in favor of introducing uncertainty in patent rights.

40. To be precise, introducing some version of the defense is optimal, and introducing the full-fledged defense is optimal up to a first-order approximation, which will be good if the defense does not represent a large, discrete change in the expected rewards to innovators who are creating significant inventor-specific spillovers.

41. The social versus private incentive to rely on trade secrets versus patents depends heavily on the importance of the disclosure function of patents. Denicolò and Franzoni (2004) study the choice between trade-secret and patent protection in a model where a second party can patent an invention that is known to have been invented previously but kept secret. Anton and Yao (2004) study the choice between patent and trade-secret protection.

42. The unadorned defense would not necessarily solve the problem of patent ambush in the standard-setting context, however. If a product standard is adopted that uses a certain patented technology before the technology has been disclosed by the patent holder, the industry member(s) that developed and contributed that technology to the industry standard would presumably qualify for the defense. Other industry members would not, however, unless the defense were extended to include parties to whom the independent inventor provided the technology before it was disclosed by the (subsequent) patentee.

43. U.S. patent law provides some very limited protections for independent second inventors, in the form of "intervening rights" in situations where a patent holder broadens its claims in a continuation patent. Even with the defense in place, however, it would still be important not to allow a vague and broad description of the invention provided in the initial application to capture subsequently introduced products that were not anticipated by the patent applicant at the time of the application was filed as those products may not be covered by the defense.

44. See Armond (2003) for an extensive discussion of these legal doctrines. She also notes that independent invention can be a defense against a finding of willful infringement under current patent law.

45. If Party B had publicized its invention, that publication would count as prior art, so Party A could not obtain a patent. However, because Party B kept its invention secret, it becomes "secret prior art" that does *not* count as prior art for the purpose of evaluating the novelty of Party A's invention. Under U.S. patent law, a second inventor can obtain a patent if the first inventor abandoned, suppressed, or concealed its invention.

46. European law provides far more generous prior user rights to party B under this fact pattern.

47. Harriel (1996) discusses the history of prior user rights under U.S. patent law and reports (563) that prior user rights can be found in almost every industrialized country other than the United States.

48. Shapiro (2006b) abstracts from the precise timing of invention between the two inventors, so the prior user rights he considers apply to second inventors who discover or use the invention prior to its disclosure by the first inventor.

49. If Party B did not conceal its invention, Party A's patent would be invalidated by prior art.

50. These latter efforts would be akin to the expenditures that are made by firms attempting to invent around an issued patent; these efforts are one form of deadweight loss associated with patents. Leibowitz (2002, 2,277) argues that the full-fledged version of the defense would encourage parties to independently reinvent patented inventions if the patent

holder seeks too high a royalty. He thus suggests a system under the patent disclosure would consist of two parts, a functional overview and a technical specification. A party would not be precluded from invoking the defense simply because it had seen the overview, which would be publicly available. In contrast, a party who had requested copies of the technical specification would be precluded from invoking the defense.

51. See U.S. Patent and Trademark Office (2006, 121), table 1.

52. The PTO recently started a peer-to-patent pilot project in which some patent applications in software design will be posted on the Internet so examiners can benefit from public comments identifying possible prior art. See "Open Call From the Patent Office," by Alan Sipress, *Washington Post*, March 5, 2007. In August 2006, the PTO established an accelerated review option under which patent applications are processed more rapidly if the applicant agrees to conduct its own search for prior art; see http://www.uspto.gov/web/offices/com/sol/notices/71fr36323.htm.

53. See Graham et al. (2003), a highly informative empirical study comparing reexaminations in the U.S. with oppositions in Europe.

54. See Federal Trade Commission (2003, 28), chapter 1, and the National Academies of Science (2004, 96).

55. Graham et al. (2003)present powerful evidence that the European opposition system is far more effective than the U.S. reexamination system. They find that the rate of opposition at the European Patent Office is more than thirty times higher than the rate of reexamination at the U.S. PTO. Furthermore, oppositions lead to the patent being revoked about one-third of the time and narrowed about one-third of the time, whereas reexaminations result in the patent being revoked only about 10 percent of the time. Levin and Levin (2003) see significant welfare benefits from introducing a patent opposition process in the United States.

56. According to the National Academies of Science (2004, 95–96), patent litigation typically does not occur until seven to ten years after the patent is issued, after which it takes two to three more years to be resolved. In January 2007, in the *Medimmune v. Genentech* case, the Supreme Court gave technology users much greater ability to control the timing of patent litigation by initiating a declaratory-judgment action to establish that the patent is invalid or not infringed.

57. Let q be the strength of the patent if it survives the reexamination. Because the patent's initial strength is 0.3, we must have $0.3 = 0.5 \cdot q + 0.5 \cdot 0$, which gives $q = 0.6$. Formally, reexamination corresponds to a mean-preserving spread in the patent strength.

58. For example, the Electronic Frontier Foundation has a "patent-busting project" (www.eff.org/patent) that seeks to overturn some Internet and software-related patents by gathering prior art and requesting PTO reexamination.

59. See *eBay, Inc. v. MercExchange*, 126 S.Ct.1837 (2006).

60. The *eBay* case is relevant to the broader question of whether "property" rules or "liability" rules are more efficient in patent law. See Calabresi and Melamed (1972), Kaplow and Shavell (1996), and Lemley and Weiser (2007).

61. See "MP3 Patents in Upheaval after Verdict," by Saul Hansell, *New York Times*, February 23, 2007. This case also illustrates the importance of complementary innovation and royalty stacking as a large number of separate parties own patents that read on the MP3

standard. See "Patent Fights Are a Legacy of MP3's Tangled Origins," by Douglas Heingartner, *New York Times*, March 5, 2007.

62. The Court explained that it would be "very grave error" to "instruct a jury that as to the measure of damages the same rule is to govern, whether the patent covers an entire machine or an improvement on a machine." *Seymore v. McCormick*, 57 U.S. 480, 491 (1853). See also *Westinghouse Elec. & Mfg. Co. v. Wagner Elec. Mfg. Co.*, 225 U.S. 604, 615 (1912: "[The] invention may have been used in combination with valuable patents made, or other patents appropriated by the infringer, and each may have jointly, but unequally, contributed to the profits. In such case, if plaintiff's patent only created a part of the profits, he is only entitled to recover that part of the net gains."

63. They discuss at length one aspect of the current doctrine, under which a party accused of infringement can avoid a finding of willfulness by obtaining a valid legal opinion that the patent is invalid or not infringed.

References

Anton, James, Hillary Greene, and Dennis Yao. 2006. Policy implications of weak patent rights. In *Innovation policy and the economy*, ed. Adam Jaffe, Josh Lerner, and Scott Stern, 1–26. Chicago: University of Chicago Press.

Anton, James, and Dennis Yao. 2004. Little patents and big secrets: Managing intellectual property. *Rand Journal of Economics* 35:1–22.

Armond, Michelle. 2003. Introducing the defense of independent invention to motions for preliminary injunctions in patent infringement lawsuits. *California Law Review* 91:117–62.

Ayres, Ian, and Paul Klemperer. 1999. Limiting patentees' market power without reducing innovation incentives: The perverse benefits of uncertainty and non-injunctive remedies. *Michigan Law Review* 97:985–1033.

Bessen, James, and Eric Maskin. 2006. Sequential innovation, patents, and imitation. http://www.sss.ias.edu/publications/papers/econpaper25.pdf.

Calabresi, Guido, and A. Douglas Melamed. 1972. Property rules, liability rules, and inalienability: One view of the cathedral. *Harvard Law Review* 85:1089–1128.

Cohen, Linda, and Jun Ishii. 2005. Competition, innovation and racing for priority at the U.S. Patent and Trademark Office. AEI-Brookings Joint Center for Regulatory Studies. http://ssrn.com/abstract=826504.

Dasgupta, Partha, and Joseph Stiglitz. 1980. Uncertainty, industrial structure, and the speed of R&D. *Bell Journal of Economics* 11:1–28.

Denicolò, Vincenzo, and Luigi Franzoni. 2004. Patents, secrets, and the first-inventor defense. *Journal of Economics Management and Strategy* 13:517–38.

Farrell, Joseph, John Hayes, Carl Shapiro, and Theresa Sullivan. 2007. Standard setting, patents and holdup, *Antitrust Law Journal*, forthcoming. http://faculty.haas.berkeley.edu/shapiro/standards2007.

Farrell, Joseph, and Robert Merges. 2004. Incentive to challenge and defend patents: Why litigation won't reliably fix patent office errors and why administrative patent review might help. *Berkeley Technology Law Journal* 19:943–70.

Farrell, Joseph, and Carl Shapiro. 2008. How strong are weak patents? *American Economic Review*, forthcoming. http://faculty.haas.berkeley.edu/shapiro/weak.pdf.

Federal Trade Commission (FTC). 2003. *To promote innovation: The proper balance between competition and patent law and policy*, October. Washington, DC: FTC.

Gallini, Nancy. 1992. Patent policy and costly imitation. *Rand Journal of Economics* 23:52–63.

Gallini, Nancy, and Suzanne Scotchmer. 2002. Intellectual property: When is it the best incentive system? In *Innovation policy and the economy*, ed. Adam Jaffe, Josh Lerner, and Scott Stern, 51–77. Chicago: University of Chicago Press.

Gilbert, Richard, and Michael Katz. 2007. Efficient division of profits for complex innovations. University of California at Berkeley. Unpublished Manuscript.

Gilbert, Richard, and Carl Shapiro. 1990. Optimal patent length and breadth. *Rand Journal of Economics* 21:106–12.

Graham, Stuart, Bronwyn Hall, Dietmar Harhoff, and David Mowery. 2003. Patent quality control: A comparison of U.S. patent re-examinations and European patent oppositions. In *Patents in the knowledge-based economy*, ed. Wesley Cohen and Stephen Merrill, 74–119. Washington, DC: National Research Council of the National Academies of Science.

Graham, Stuart, and David Mowery. 2004. Submarines in software: Continuations in U.S. software patenting in the 1980s and 1990s. *Economics of Innovation and New Technology* 13:443–56.

Griliches, Zvi. 1979. Issues in assessing the contribution of R&D to productivity growth. *Bell Journal of Economics* 10:92–116.

Hall, Bronwyn. 1996. The private and social returns to research and development: What have we learned? In *Technology, R&D and the economy*, ed. B. Smith and C. Barfield, 140–61. Washington, DC: Brookings Institution and American Enterprise Institute.

Harriel, Kyla. 1996. Prior user rights in a first-to-invent patent system: Why not? *IDEA: The Journal of Law and Technology* 3:543–50.

Hegde, Deepak, David Mowery, and Stuart Graham. 2007. Pioneers, submariners, or thickets: Which firms use continuations in patenting and why? NBER Working Paper no. W13153. Cambridge, MA: National Bureau of Economic Research.

Henry, Emeric. 2007. Runner-up patents: Is monopoly inevitable? London Business School, April. http://papers.ssrn.com/sol13/papers.cfm?abstract_id=922316.

Hopenhayn, Hugo, Gerard Llobet, and Mathew Mitchell. 2006. Rewarding sequential innovators: Prizes, patents, and buyouts. *Journal of Political Economy* 114:1041–68.

Jaffe, Adam, and Josh Lerner. 2004. Innovation and its discontents: How our broken patent system is endangering innovation and progress, and what to do about it. Princeton, NJ: Princeton University Press.

Kaplow, Louis, and Steven Shavell. 1996. Property rights vs. liability rules: An economic analysis. *Harvard Law Review* 109:713–90.

Kitch, Edmund. 1977. The nature and function of the patent system. *Journal of Law and Economics* 20:265–90.

Klemperer, Paul. 1990. How broad should the scope of patent protection be? *Rand Journal of Economics* 21:113–30.

La Manna, Manfredi, Ross MacLeod, and David de Meza. 1989. The case for permissive patents. *European Economic Review* 33:1427–45.

Lamb, David, and Susan Easton. 1984. *Multiple discovery: The pattern of scientific progress.* Trowbridge, UK: Avebury.

Leibowitz, John. 2002. Inventing a nonexclusive patent system. *Yale Law Journal* 111 (8): 2251–87.

Lemley, Mark. 2001. Rational ignorance at the patent office. *Northwestern University Law Review* 95:1497–1532.

———. 2007. Should patent infringement require proof of copying? *Michigan Law Review* 105:1525–36.

Lemley, Mark, and Kimberly Moore. 2004. Ending abuse of patent continuations. *Boston University Law Review* 84:63–123.

Lemley, Mark, and Carl Shapiro. 2005. Probabilistic patents. *Journal of Economic Perspectives* 19:75–98.

———. 2007. Patent hold-up and royalty stacking. *Texas Law Review,* forthcoming. http://faculty.haas.berkeley.edu/shapiro/stacking.pdf.

Lemley, Mark, and Ragesh Tangri. 2003. Ending patent law's willfulness game. *Berkeley Technology Law Journal* 18:1085.

Lemley, Mark, and Philip Weiser. 2007. Should property or liability rules govern information? *Texas Law Review* 85:783–841.

Levin, Jonathan, and Richard Levin. 2003. Benefits and costs of an opposition process. In *Patents in the knowledge-based economy,* ed. Wesley Cohen and Stephen Merrill, 120–36. Washington, DC: National Research Council of the National Academies of Science.

Mansfield, Edwin, John Rapoport, Anthony Romeo, Samuel Wagner, and George Beardsley. 1977. Social and private rates of return from industrial innovations. *Quarterly Journal of Economics* 91:221–41.

Maurer, Stephen, and Suzanne Scotchmer. 2002. The independent invention defense in intellectual property. *Economica* 69:535–47.

Merton, Robert. 1957. Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review* 22:635–59.

Miller, Joseph Scott. 2004. Building a better bounty: Litigation-stage rewards for defeating patents. *Berkeley Technology Law Journal* 19:667–739.

National Academies of Science (NAS). 2004. *A patent system for the twenty-first century.* Washington, DC: The National Academies Press.

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

Shapiro, Carl. 2006a. Injunctions hold-up, and patent royalties. http://faculty.haas.berkeley.edu/shapiro/royalties.pdf.

———. 2006b. Prior user rights. *American Economic Review Papers and Proceedings* 96 (2): 92–96.

Tandon, Pankaj. 1982. Optimal patents with compulsory licensing. *Journal of Political Economy* 90:470–86.

———. 1983. Rivalry and the excessive allocation of resources to research. *Bell Journal of Economics* 14:152–65.

U.S. Patent and Trademark Office. 2006. *Performance and accountability report*. http://www.uspto.gov/web/offices/com/annual/2006/index.html.

Vermont, Samson. 2006. Independent invention as a defense to patent infringement. *Michigan Law Review* 105:475–504.