

Raising the Bar of Patentability: Eliminating the Patent Disincentive

Jason Rhineland

27 April 2010

I. Introduction

The fundamental task of the patent system is to encourage innovation. Generally, through offering an exclusive right over an innovation, firms can recoup the high fixed cost of invention through the higher price possible with the ability to exclude others from the innovation. At a basic level, this seems intuitive: provide innovators with the necessary rewards to secure innovation. Despite this, evidence shows the existence of a patent paradox: patents are still obtained whose expected returns fall below the cost of obtaining the patents. Though various explanations have been proposed that attempt “solve” the paradox—some more compelling than others—its existence nonetheless suggests that patents must be serving as a drag on innovation—certainly at the aggregate level, if not at the individual level. This can be explained, at least partly, as a patent prisoner’s dilemma: firms are better off obtaining patents than not, but would be better off still if no one obtained patents.

This effect is not, however, uniform across industries currently subject to the patent system; the reduction of patent returns seems largely caused by the increased litigation costs resulting from uncertainty in industries where the width of patent claims and patent validity is much larger. Where infringement is easy to determine and detect, such as in the pharmaceutical and chemical industries, patents—or, at least, portfolios of related patents—appear to have positive returns at both the individual and aggregate levels. Once the increased cost of litigation to determine and deter infringement is factored in, the returns become negative. However, opting out of the patent system is not an option: an individual firm will be subject to infringement of other firms’ patents, whether or not the firm engages in the patent system itself, and thus holding patents may provide a stronger bargaining position should infringement, intentional or otherwise, occur.

Amendments to the patent system are needed so as to correct innovation disincentives present without significantly affecting the incentives provided by patents where such incentives are positive. This paper proposes a narrowing of patent validity to reject patents with negative incentives while not significantly affecting patents on well-defined innovation.

II. The existing patent system

Under patents, as currently implemented, we have socially and historically imposed requirements for patentability of novelty and non-obviousness¹. In particular, the novelty and non-obviousness requirements together are intended to constitute the breadth of a patent—what the patent covers, how the patent

¹ “Non-obviousness” is sometimes instead referred to as a required “inventive step;” regardless, the essence of the requirement is the same.

contributes a new invention, and, in doing so, defines the contours of the patent within which other inventions may infringe the patent. These controls thus allow manipulation of the incentives to invent through changing the value of a patent, and through patent disclosure, deliver a social benefit of dissemination of patented inventions. The precise boundaries, however, of “novel” and “non-obvious” are relatively difficult to determine in practice, tending to be defined by legal rulings of patent litigation.

There are two opposing components to the social value of a given invention: on the one hand, a patented innovation requires disclosure of the details of the innovation. For innovations that would otherwise not be disclosed or independently invented—i.e. trade secrets—this is a clear increase in social value. On the other hand, for patented inventions that would be publically released in some form², we need to consider the value of the patent disclosure against the information determinable from publically available implementations of the innovation itself. If the disclosure benefit of the patent information is low relative to the disclosure benefit of the invention itself—perhaps because the innovation can be easily “backwards-engineered³,” or because the implementation details of the innovation itself are obvious⁴—then the social benefit of a given invention without patent exceeds that of the same invention with a patent. Though this, of course, ignores the incentive to research in the first place, it suggests that there can exist patents with negative individual and social values compared to non-patentability of the same good. As the cost of obtaining and maintaining a patent increases, or the probability of court-confirmed validity and enforcement decreases, the value of a patent also decreases, and potentially becomes negative. For efficiency, such patent should not be pursued, yet this is not the case. If characterized by particular industries, such pursuit would indicate that patents provide a *disincentive* for creation in those industries. Why patents with an *ex ante* negative expected return might be obtained is considered later in this section.

In the case of pharmaceutical and chemical companies, clearer benefits to patentability exist. Patents can be reasonably and concisely defined on a particular drug or compound, providing the positive expected returns to research. Fundamentally, identification of infringement here is relatively easy. This has two effects: on the one hand, it lowers the cost of demonstrating infringement, while also potentially lowering the probability of a patent being overturned. More clearly defined patent boundaries, combined with the lower expected litigation costs, thus keep expected patent maintenance costs low, while the patent delivers potentially large payoffs for exclusive control of the patented substance through the patent’s

² Depending on the patent, perhaps the innovation itself, or some good making use of the patent.

³ For example, through a chemical analysis which can determine the makeup of a particular compound, or the “reverse compiling” of a piece of software to reveal its implementation.

⁴ Of course, such an invention would still need to pass an overall requirement of non-obviousness, but the obviousness of the steps need not necessarily imply the obviousness of the invention itself.

lifespan. The relatively low costs, and high individual values, suggest that the patent system provides positive incentives for these types of firms.

Similar patent value exists in any industries that exhibit the same fundamental characteristics. First, a patent must deliver significantly large individual patent value⁵. Second, identifying infringement with a relatively low rate of error—either error in finding infringement when it should not be found, or in not finding infringement when it should⁶—is required to avoid diminishing the expected value of the patent. Third, the litigation costs in maintaining and defending a patent must be relatively low. Finally, patent boundaries must be clearly definable and easily identifiable, both to avoid unintentional and unexpected infringement⁷.

The opposite end of the spectrum of patentability occurs in software and algorithm patents, where claims tend to be widely and inconsistently interpreted and the outcomes of infringement cases are often highly unpredictable. As Bessen and Meurer (2008) note, such patents are, at their core, patents on abstract ideas. The difficulty that emerges is how to interpret the width of a claim on an abstract construction. Unlike a patent on a chemical process, a software patent imposes significantly increased costs of accurate infringement determination. While individual patent valuations vary, there is some indication that they are not especially high, and certainly much less on an individual, single patent level, than a chemical process patent⁸. Evidence furthermore suggests that litigation costs for these types of patents is high, due in part to the increased costs of determining infringement. Finally, because of their abstract nature, software and algorithm patents are rarely clearly defined explicitly, while widely ranging court decisions regarding software patents do not provide sufficiently clear implicit boundaries. Evidence suggests that software patents, in general, have an overall negative expected value once accounting for legal expenses⁹, but as discussed later, positive

⁵ As discussed later, the “individual” patent value need not be derived from the patent alone—it needs only to provide more value, perhaps in its contribution to a patent portfolio, than it costs to obtain.

⁶ This assumes, of course, some notion of a “correct” infringement ruling exists; when patent contours are defined by patent common law, every standing court ruling regarding patents must, by definition, be “correct.” While I would prefer to take the broader view that correctness is derived from the point of view of economic efficiency and patent optimality, rather than the strictly legal interpretation, ultimately patent decisions are derived from firms’ beliefs of patent infringement as determined by courts. Thus it might be more useful to defined correctness in terms of consistency and predictability of court rulings, rather than economic efficiency, so that patents’ value can be accurately predicted *ex ante*.

⁷ Unexpected infringement being that where the infringer believed itself to not be infringing, and unintentional infringement being where the infringer was unaware of the existence of the patent.

⁸ Bessen and Meurer look at these disparities in value in significant detail.

⁹ Barring a very small number of spectacular exceptions, such as NTP, Inc.’s patent infringement case and settlement against BlackBerry creator Research In Motion.

value may emerge from maintaining a cluster of related patents usable against would-be infringement claimants.

A similar story exists in many industries connected with, but outside, the software industry. Bessen and Hunt (2004), in a study of software patents, found that only around 5% of software patents come from inside the software industry itself; most, they found, came from related fields—semiconductor firms, computer manufacturers, etc. This should be at least somewhat expected: hardware designs often incorporates algorithms or functionality that, while encoded physically instead of virtually, can nonetheless infringes upon a software or algorithm patent. This is not to suggest that this is universally the case for related firms: in many cases electronics patents may well be more clearly defined, much less abstract, and more valuable than software patents. To the extent that electronics firms cross over into patents of similar abstractness as those covering software and algorithms, however, it is expected that the story for software patents is largely the same: patents for which it is difficult to accurately determine infringement, and which may similarly have negative individual returns. Also like software patents, it may be the case that relatively abstract patents are again clustered to provide a defence against potential infringement.

Bessen and Meurer (2008) explore the costs and profits associated with patents in the United States and find startling evidence that suggests that there is indeed much higher spending on patent litigation than profits from patents for patenting firms outside the chemical and pharmaceutical industries, particular in the software industry. Combined with patent filing and associated attorney fees, they conclude, patents are thus providing a disincentive for innovation rather than the intended incentive. They attribute much of this large cost to the uncertainty surrounding patents, citing numerous court cases where the patentee has interpreted patent claims in novel, non-obvious ways, which have been upheld by both patent offices and patent courts¹⁰.

There exists already a large class of innovations that are not patentable: namely, any obvious or non-novel innovation. For reasonable definitions of obviousness and novelty, these exclusions make sense: an obvious patent is, by definition, obvious; the fact that it is not been previously invented could well be a result of a lack of application for the innovation rather than a lack of consideration of the innovation. Introducing patentability into this class of invention would necessarily hinder innovation: the risk of infringement, and potential liability, would be large. The requirement of clearing patent rights, so as to be able to creation without fear of innovation, would impose a significant cost on any innovation, including that which is currently entirely outside the patent regime. Such a system would only serve to raise barriers to entry for innovation for all but the largest

¹⁰ Examples abound; for instance US Patent No. 4,528,643, issued to firm E-Data, a “System for Reproducing Information in Material Objects at a Point of Sale Location”, covered reproducing digital tapes at kiosks, but was later used to claim infringement by large numbers of electronic commerce firms. Also see later discussion of RIM and NTP.

firms, who would still be able to innovate (even at a small level) by extracting cross-licensing agreements with other large firms through asserting their own patent portfolio thicket. Gains to large firms thus exist not in persuading innovation to better compete with other firms, but in joining an exclusive club of concentrated market power. A minimum requirement of patentability is thus not controversial: excluding frivolous and obvious innovations from patentability is clearly required for efficiency.

Though pharmaceuticals and related industries were identified above as benefitting from patents, it may not always be the case that all developments within those industries fit within the clear benefit category. Heller and Eisenberg (1998) explore the role of patents at deterring innovation in the related biomedical industry, identifying a “tragedy of the anticommons.” In 1980, the United States Congress attempted to encourage private investment and the transfer of research from “upstream” researchers—universities and similar institutions—to commercial, “downstream” firms, by altering patent rules to allow patents on publically funded research discoveries, which had previously been unpatentable. The result was a change from downstream firms being able to freely use the public domain results of public research to requiring negotiation with the patent holders of the discoveries. The inherent overlap in the property rights, and the associated transaction costs in negotiating clearing rights—particularly for patents over DNA sequences not necessarily related to a particular biological function or commercial product—dampens research to a level, Heller and Eisenberg argue, that risks creating a tragedy of the anticommons resulting in fewer downstream use of the patents.

The disincentive to innovation results, in part, from the “tragedy of the anticommons,” a term introduced by Heller (1998) in describing property rights in former socialist countries, and shortly thereafter extended to patent rights by Heller and Eisenberg. The argument is basically a transaction cost story: when rights significantly overlap, because of many overlapping patents¹¹, the transaction costs of securing the rights to create a product using the patented innovations also increases. As increased patent filings increase the number of overlapping patents covering a particular innovation, the incentive to innovate is reduced in two ways: firstly, it increases transaction costs for downstream innovators, resulting in fewer downstream innovations and thus lower royalties for upstream innovators. Secondly, the larger number of patents covering a particular innovation results in a lessening of possible royalties associated with individual patents.

Hunt’s (2006) construction of a model of firms’ choices to patent and innovate, deliberately treating the decisions between the two as independent, sheds more light on the patent anticommons. In his model, so long as there was little patent overlap, optimizing firms treated patents as complements to research and development, as would be expected. Hunt’s model become much more interesting as patent overlapping was increased: for sufficiently high levels of patent overlap,

¹¹ In Heller’s socialist country examination, the problem existed in overlapping property rights arbitrarily and excessively granted by the newly non-socialist countries.

firms—whose incentives are determined not only by the rents available through a firm’s own patents, but also the rents available from the infringement of rival firms upon the firm’s patents—optimally treat patents and R&D as substitutes rather than complements. Thus, at least in his model, Hunt established a correlation between patent overlap and decreased R&D in agreement with Heller and Eisenberg’s anticommons argument¹².

If patents are, in fact, providing negative expected returns, one must consider the “patent paradox:” why such patents are obtained in the first place. F. M. Scherer (2001) argues that the distribution of awards tends to turn research and development into a sort of lottery, with very large payoffs go to very few patent holders, while most patent holders receive little. The argument is interesting, though not particularly compelling as it depends, as Scherer points out, upon risk-seeking behaviour, at least locally, among individual patent holders, particularly given the size of negative patent returns identified by Bessen and Meurer.

The lottery idea can be extended to pertain not only to patents but to the research and development process overall. Given research and development, several results are possible: the research might fail, that is, lead to nothing of particular value. The research might lead to a poor-value patent, which might be obtained simply to “troll” other firms that develop a similar idea. Research might lead to an innovation of substantial value, at which point the firm can decide to patent—thereby disclosing the innovation but receiving legal protection—or can decide to keep the innovation as a trade secret.

Another extension of the patent lottery exists in the form of the aforementioned patent “trolls”: that is, firms who produce nothing covered by the patent in question, but seek to extract rents from other firms after other firms have clearly and successfully infringed¹³. Economics and legal literature is rife with examples of such behaviour, which tend to stem from an incongruity in patent claim interpretations at the time of filing and at the time of a charge of infringement. Indeed the optimal strategy for the patent troll is to seek claims that can be interpreted sufficiently narrowly at the time of filing to be accepted, but sufficiently widely at the time of infringement so as to support a threat of infringement litigation. Moreover, the successful patent troll has an incentive to obfuscate patent claims so as to make the patent less likely to be found by the potentially infringing firm who might otherwise navigate around the patent. Note, however, that patent trolling is not necessarily done with *ex ante* motivation: it is quite possible for a misvalued or failed patent idea, which was patented for one of the other reasons discussed in this paper, to become the object of patent infringement lawsuits,

¹² It is worth noting that, while Hunt’s model demonstrated a correlation, the patent overlap parameter was exogenous. An interesting extension of Hunt’s model would be to allow firms to also optimize through a choice of patent overlap, perhaps with a constraint on the maximum permissible overlap (thus simulating the “inventive step” of obtaining a patent).

¹³ See *supra* note 10.

perhaps under altered patent interpretation¹⁴. It can also at times be difficult to draw a line between legitimate innovation and patent infringement and the patent trolling described here. Whether patent trolling occurs *ex ante* or *ex post*, its existence—a symptom of permitting abstract or indefinite claims—nonetheless harms innovation by requiring firms to anticipate a possible patent infringement case by a patent troll.

Another theory of the patent paradox, put forward by Richard Levin (1986), presents some evidence of intra-firm signalling as a possible paradox explanation: individual researcher productivity is signalled to employers by submitting patents, essentially using the patent office as an impartial arbiter of an employee's productivity. The explanation can be supported by assuming an expectation of shirking responsibility if left unmonitored, and the difficulty in assessing employee productivity by managers, especially in research and development. Wagner and Parchomovsky (2005) attempt to dismiss the explanation on the grounds that it falls victim to the paradox itself: firms should not want a patent that has a negative expected return, and thus using the patent office in this way should be avoided. This dismantling begs the question, however, by ignoring the value of the signal being provided by the patent office: if a firm believes that using the patent office for internal productivity signalling is sufficiently valuable, returns after the patent are granted are unnecessary. That said, the relatively high filing costs of patents makes this explanation largely implausible on its own: one would expect firms to be able to determine productivity more accurately and at less cost through other means. The explanation may, then, help explain the paradox in marginal cases: where the expected return from a patent is slightly negative, the expected return *including* the signalling value may become positive. While intra-firm signalling may make up part of the paradox story, it seems unlikely that it solves more than a small part of the patent paradox.

A more compelling explanation of the patent paradox is present in Wagner and Parchomovsky (2005), who propose patent portfolios as a plausible patent paradox explanation. The authors suggest a resolution of the paradox by arguing that the relevant value of a patent lies not in the individual patent, but in the effect on the aggregate value of a collection of related patents. Benefits thus emerge not from the advantages provided by one patent, but the advantages provided by a cluster of related patents; the marginal value of the patent is not simply the value of the patent itself, but the increased aggregate value of the firm's patent portfolio. Within the framework of a portfolio, individual patents can cover individual aspects of a particular type of technology or development that cannot be effectively protected under a single patent. The portfolios in effect form "super-patents:"

¹⁴ In addition, where prior art invalidates (or causes a lack of acceptance of) a patent, *ex ante* trolling is assumed to be rather difficult: one must anticipate a potential avenue of future. This suggests that *ex post* trolling, primarily through reinterpretation of the claims of existing, accepted patents in the face of commercially successful innovation is likely a much larger source of patent trolling.

protection much wider than that offered by patents by joining together patents related to different aspects of a technology¹⁵. Among a number of advantages identified by the authors is the notion that the broader coverage of a super-patent confers increased predictability of legal outcome, and the associated lowering of legal costs in defending and affirming a patent right—even if one patent’s validity is in question, the likelihood of *all* patents related to a particular super-patented invention being overturned is significantly reduced¹⁶.

Patent portfolios may thus act as a sort of self-insurance, mitigating the risk of losing patents, regardless of whether the patent revocation is due to legitimate or erroneous patent office or court rulings. Even with a relatively high rate of overturning patents, a cluster of a significant number of patents allows the firm to pool risk, significantly reducing the probability of dismantling the super-patent, thereby increasing the expected value of the super-patent by increasing the probability of its joint-validity, rather than affecting specific patent returns.

An alternative explanation for at least part of the patent paradox comes in the form of defensive patent portfolio races, as proposed by Hall and Ziedonis (2001). Obtaining patents may operate as a sort of arms race, where patent development occurs not to exclude others from pursuing a development, but to protect a firm from being excluded by the patents of other firms and to give the firm increased negotiation power. By obtaining a patent (or patent portfolio), the firm secures an ability to negotiate with potential challengers and infringers, which can potentially lead to explicit or implicit cross-licensing agreements¹⁷. The explanatory power of the patent arms race is significantly strengthened when combined with the idea of patent portfolios; the defensive ability of a patent portfolio is increased over a patent for much the same reasons that the offensive abilities of a portfolio are increased. If patents indeed function in this way, the decision to patent is based not only on the return of the patent, but also on the extent to which a defensive patent

¹⁵ Wagner and Parchomovsky sensibly explicitly exclude *unrelated* patents from their definition of patent portfolios: indeed the argument becomes less credible for a collection of unrelated patents, where additional patents have little effect on the other, unrelated patents.

¹⁶ Patent overturning need not be particularly error-free when one has a patent portfolio, and as such may protect against erroneous or unpredictable patent determination.

¹⁷ Implicit in the sense that such agreements need not be stated because firms do not enforce their patents for fear of retaliation. Suppose, for instance, two firms each a collection of patents necessary to produce a good; each could prevent the other from producing, but may prefer duopoly competition to an alternative with no good at all. Retaliation to super-patent infringement lawsuits, often as a form of negotiation, is quite common; see, for example, Nokia suing Apple over 10 of Nokia’s communication-related patents (<http://www.nytimes.com/2009/10/23/technology/companies/23nokia.html>), and the countersuit by Apple over 13 of Apple’s patents related to the design of the Apple iPhone (<http://www.nytimes.com/2009/12/12/technology/12apple.html>).

can limit potential infringement losses. Thus a patent can appear to have a negative return, though actually offers positive incentive over not having to patent¹⁸.

If indeed patent portfolios and defensive patents can explain the patent paradox it must necessarily be asked what implications a *de facto* requirement of patent portfolios for innovation imposes. On a basic level, individual patent values that provide a disincentive to patent cannot encourage innovation; patentable but unpatented innovation is provided no incentive by a patent system, but in fact has a disincentive present by the risk of potential infringement. Securing a patent or patent portfolio can mitigate such risk, but this only serves to impose higher cost on the innovation, also providing a disincentive for innovation¹⁹. Under the defensive portfolio patent paradox explanation, the pursuit of patents is a drag on innovation; the existence of a patent system used defensively discourages innovation.

Patents cannot be examined in absence of the associated solution to dealing with patent conflicts: licensing and associated royalties. Through licensing we allow division between and specialization among upstream research firms and downstream production firms making use of a patented innovation. Allowing such behaviour is, of course, beneficial: whatever the allocation of the patent right to an innovation, gains to efficiency can occur through the ability to trade that right. It is important to note, however, that the efficiency gains from licensing are offset by the previously discussed anticommons effect induced by higher transaction costs as patent concentration increases.

While licensing of an individual patent may result in efficiency gains, it is not necessarily the case that the gains hold in aggregate, especially once patent cross-licensing is considered. In some cases, cross-licensing is clearly beneficial, where it allows innovation and production where it otherwise could not occur. For example, three firms, each holding overlapping portions of what would be a “super-patent” portfolio if held by a single firm, are each unable to produce at all without securing a license from the other two firms. A cross-licensing agreement between the three firms, on the other hand, allows the three firms to compete in production or use of the innovation. As far as production of the good is socially beneficial, so is the benefit of cross-licensing.

Cross-licensing can, however, be manipulated for producer gain, even without increasing market power, by using royalties to jointly increase the marginal cost of the participating firms. Continuing with the example of three firms, if each firm agrees to pay the other two firms r for each unit it produces, the three firms each increase their marginal cost by $2r$. As long as r is not so large as to push the firm’s marginal cost above the profit maximizing monopoly price, such an

¹⁸ More accurately, the patent offers fewer disincentives than pursuing the innovation without the patent—but there is still a net disincentive from the existence of other patents in the class.

¹⁹ Pulling from various sources, Parchomovosky and Wagner (2005) estimate the cost of obtaining a single patent—including filing, attorney, issuing and renewal fees—at between \$10,000 and \$30,000.

agreement will increase the profits of each of the firms, even if the firms charge no more than their apparent marginal cost. Moreover such collusion is sustainable, as the alternative of cheating from the agreement is, ignoring antitrust concerns, legally enforceable patent infringement²⁰.

In an ideal, modelled world absent of transaction costs (and antitrust concerns), cross-licensing might be the end of the story; firms would costlessly negotiate the Coasian space defined by the granted patents. In reality, the transaction costs between firms are likely to be large. Negotiating patent portfolio cross-licensing between just two firms with complex patent portfolios would be difficult already; extending the negotiation to a multitude of firms of different sizes and different patent valuations operating within the same range of related patents would be substantially more difficult. The effect of patent reform on cross-licensing will be returned to later in this paper.

The primary difference between industries where patents are a net benefit to innovation and industries where patents are a net loss to innovation is the difficulty in accurately detecting infringement. For pharmaceutical industries, detection of infringement is relatively easy: a substance can be tested with some objectivity for infringing characteristics. For the opposite end of the spectrum, for those innovations whose non-patentability is due to obviousness or non-novelty, detection of infringement and validity of patents would be exceedingly difficult; in these areas we have clearly decided that patentability would cause significant harm, while providing little benefit. The in-between areas, containing abstract patents such as software patents, are characterized by a similar difficulty of determining infringement and validity; high litigation costs and unpredictable court decisions regarding upholding or overturning patents induces costly defence and a higher probability of costly litigation.

Courts can err in their patent rulings in, generally, one of two ways: infringement can be found where it should not, and infringement can fail to be found where it should be. Higher rates of either type of error introduce uncertainty into the patent system, which increases the chance of infringement litigation. As previously discussed, for abstract patents, such as software patents, accurate infringement determination—by either courts or potential innovators—is very difficult, and understandably, the cost and probability of litigation connected to a patent—and thus the expected cost of litigation—increase.

²⁰ There is some legal precedent to ignoring antitrust concerns in this case. The most notable example of such exists in the 1931 US Supreme Court case *Standard Oil Co. (Indiana) v. United States*, 33 F.2d 617 (D. Ill. 1929), rev'd, 283 U.S. 163 (1931). The case concerned related patent pooling and cross-licensing by oil companies concerned with oil cracking (breaking down crude oil into various usable compounds). The court ultimately ruled that the agreements were *not* anticompetitive as there was no explicit agreement to undertake anticompetitive practices such as price-fixing or limitations beyond the scope of the patented technologies. See also Gilbert (2009) for a brief discussion of the case. That the court erred in its ruling is a topic left to other papers.

When combined, the high patenting and expected litigation costs caused by a high likelihood of infringement lawsuits are serving as a disincentive for innovation. Patents end up failing their most basic function, effecting innovation, for a significant class of patents. Such a system is failing; patents, rather than encourage innovation, act as a barrier to legitimate innovation. The decision process of an innovator with respect to innovation and patenting is thus represented by the following figure:

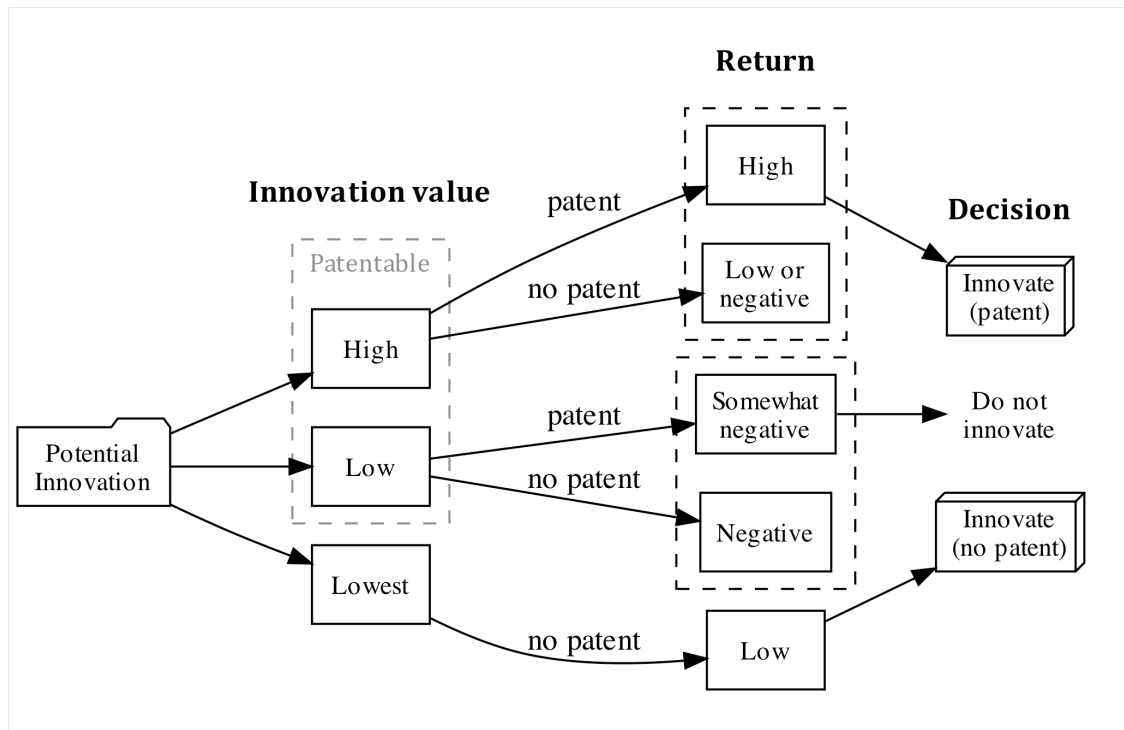


Figure 1: Innovation process with weak patents

Patenting of the low-value, but patentable, class of innovation has largely become a defensive mechanism: firms choose to patent low-value innovations not because the patent delivers positive expected returns, but rather because the expected value of the patent is less negative—by dissuading others from claiming infringement—than the alternative of non-patentability, with its larger risk of infringement.

III. Raising the bar of patentability

An alternative approach to patent granting is clearly required. In essence such a system needs to rework patent grants where innovation is currently be discouraged by either refining patent protection, or in cases where such refinement is impossible, throwing out patentability. Thus such a system needs to change the

decision process in Figure 1 into a system where the “Low” value innovators still have an incentive to innovate, as shown in the following Figure 2.

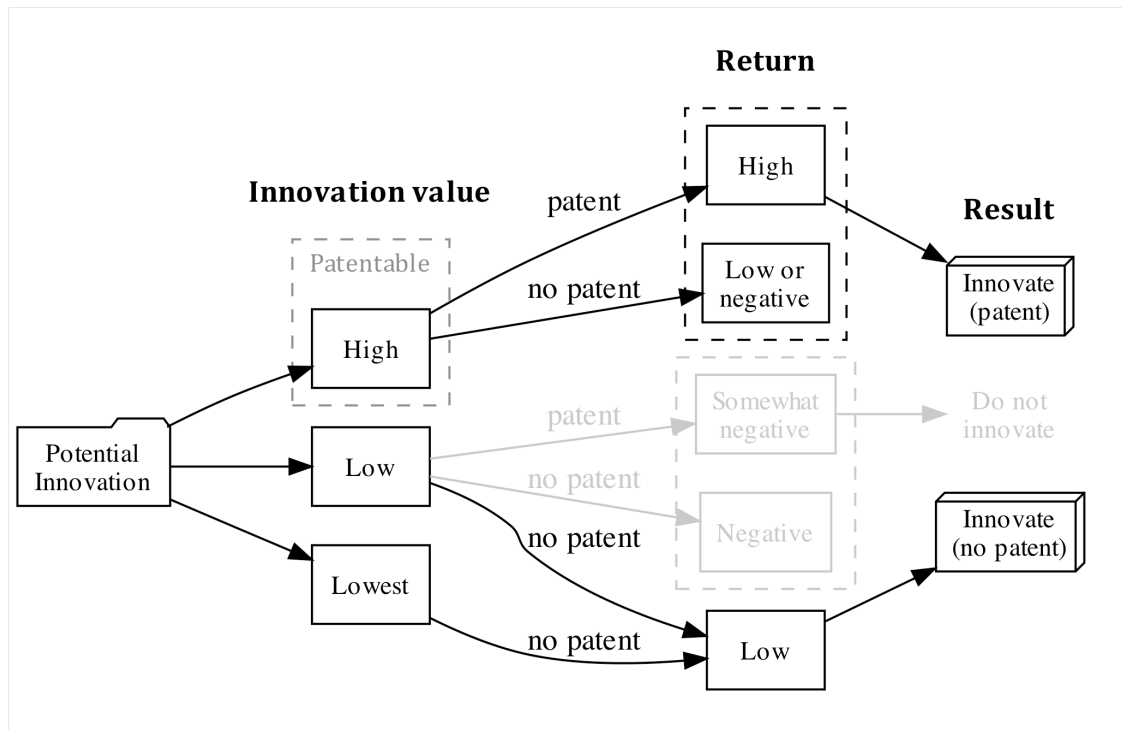


Figure 2: Innovation process without low-value innovation patents

The first step that could be taken is a strengthening of the non-obvious requirement of patents to a requirement of being “unlikely.” Specifically, patent scope needs to be redefined to exclude innovations that would likely occur without patent protection. For example, a patent was granted to Amazon for a “one-click” purchase option²¹. This “innovation” allows customers to have their ordering information stored and, when authenticated with Amazon’s website, to order books and other merchandise through a single click, as opposed to the more traditional approach of adding merchandise into a virtual shopping cart and proceeding through multiple steps to place an order. It seems highly unlikely that Amazon would not have developed the feature of their online store in the presence of non-patentability; rather than provide additional incentive for Amazon to innovate, this particular patent only served to allow Amazon to exclude others from the practice. Where innovations would occur without a patent system, patents only hold back development. In the case of a costly drug development, in contrast, the justification for patents is much stronger: the innovation would not occur without the ability to exclude others from making use of it.

An additional patent problem is in the abstractness of claims permitted by patents. Software patents in particular, but not exclusively, are necessarily abstract

²¹ US Patent No. 5,960,411

in their claims, and this abstraction leads to enormous patentability uncertainty. A revised patent system needs stricter patent requirements both at the claim level and at the interpretation level. Bessen and Meurer (2008) mention a great many patents whose claims and interpretations at the time of infringement allegation fall wildly outside the original claims: for example, U.S. Patent No. 4,528,643 originally covered the innovation of kiosks selling digital tapes at retail outlets, but the patent holder, E-Data, alleged infringement—and was confirmed by courts—on a wide range of e-commerce websites, claiming that digital content sold via websites infringed its patents. Bessen and Meurer propose as a potential solution an “indefiniteness” standard, where patents that attempt to make wide, non-specific claims would be invalidated. While such a requirement would, indeed, eliminate potential patent problems, a more fundamental problem is that patents on abstract ideas can be granted at all. Allowing abstract patents in effect allows exclusion of innovation that may not yet exist, and in any case is not a copy but a parallel innovation. Software patents, for example, typically cover not only a particular implementation—which, for non-trivial functionality, would already be protected under the copyright of the original software—but also lay claim to all implementations of the same goal as infringing behaviour. This determination, however, is difficult: comparison between algorithms is a topic in which, despite receiving a great deal of attention in fields such as computer science, determining or “proving” the equivalence of complex algorithms is often more art than science, and in any case is enormously difficult for all but the simplest algorithms.

Any patent reform must take into account the essential goal of a patent system: encouraging innovation. In industries where patents discourage innovation—software patents, business method patents, and areas currently not eligible for patent protection—they need to be (or should remain) eliminated. The mere existence of patents does not justify that existence: as previously discussed, the notion of patents and patent portfolios as a defensive arms race “solves” the patent paradox without actually providing positive, aggregate patentability benefits. Any patent system adjustment thus needs to take into account not only the individual return of a patent, but also the return under the existence of other patents in the same class, lest the result be a fallacy of composition.

By adjusting the patent system to accommodate stricter limits and throwing out patent classes that do not, in aggregate, benefit from patents, we would expect to see an encouragement of innovation. In particular, such changes would eliminate the risk of infringement by making patent claims more predictable and transparent, eliminating the chance of innovation being held up by trivial patents, and eliminate the uncertainty present in patent infringement determination. Together, these changes would make patents less costly by reducing the cost of enforcement and making infringement clearer.

The changes would also eliminate “patent trolls:” firms who patent an idea not to protect their innovation but as a lottery ticket, waiting for the idea to be used successfully by another party, and then pouncing on another firm for huge court findings or settlement offers. One of the most notorious of such cases is that

between NTP, Inc., a non-innovating patent holding firm, against RIM, creators of the BlackBerry phone; the two firms settled the patent dispute for \$612 million. What makes the case interesting, against the notion of patents-as-defence, is that, because NTP had no involvement in innovation using their patents, RIM's patents could not be used defensively²². While defensive patents and patent portfolios may be a functioning system between innovators, it provides no defence whatsoever against non-innovating patent holders seeking to extract rents from innovators.

Any changes to patent system boundaries are likely to cause a disincentive for some industries. Narrowing the allowable scope of claims in a single patent, for example, might require firms to increase the size of patent portfolio to obtain new protection, but the hope is that, with reduced scope of existing patents, patent filing costs—both for patent examiners and patent litigators—ought to be reduced, which might mitigate or even eliminate the increased cost. Some patent holders, notably those who seek broad patents on fairly abstract ideas would be entirely excluded—but as much as those patents represent rent-seeking rather than innovative behaviour, elimination of such behaviour would increase the expected returns to innovation—by decreasing potential litigation and settlement costs—and thus encourage, rather than discourage, innovation.

Innovation is further encouraged by such a change from its effective clearing of the “patent thicket”—that is, the many patents for any given innovation for which infringement is uncertain and difficult to determine—thereby reducing another impediment to innovation. Rather than be dissuaded by the possibility of an innovation being held up by unknown or unclear patents, firms are freer to innovate knowing that the class of innovation in which they operate is not subject to patent claims.

For patent cross-licensing discussed earlier, eliminating a class of patents would, in effect, be the equivalent of costlessly mandating cross-licensing among all patent participants. Rather than requiring investigation into possible infringement and negotiation of licensing agreements to avoid those infringements—and the inherent holdup problem if the clearance and negotiation occur after development of the infringing product²³—are eliminated. Similarly, the potential antitrust situation of firms colluding through patent royalties is eliminated. In situations where cross-licensing was already working efficiently to solve the patent thicket, elimination of patents will be, at least in aggregate, no less efficient than the negotiated solution.

²² Distinguishing between a troll and a pure research firm is, as previously mentioned, difficult, but the NTP case seems to be regarded as a clear example of a patent troll: NTP's patent was relatively abstract, and RIM's infringement appears to have been as a result of parallel, independent development of the same idea, rather than intentional infringement or copying of the original patentee's innovation.

²³ Perhaps due to uncertain *ex ante* product implementation details, or patents granted after an initial patent clearing effort in first-to-file patent jurisdictions.

For the previously identified “clear-benefit” industries, where patents have sufficient individual value, and are clearly identifiable, such a change should have little effect²⁴. For the most part, patents for pharmaceuticals would be unaffected, though some patents covering overly-broad, abstract claims on areas such as gene patents might be affected²⁵—but given the likely accelerating importance and research in this field, the lack of patentability over the generic, abstract aspects of the research ought to similarly foster, rather than prevent, the aggregate level of innovation.

Essentially, the patentability question centres on the position of the patentability requirement bar. By allowing abstract patents, innovation in classes covered by those patents is discouraged. In its characteristics, the class of patents that would be affected by a raised patentability bar has more in common with the current non-patentable innovation class than with classes of patents with much clearer benefits. The remedy, similarly, is no different from the (existing) solution for obvious or non-novel innovation: innovation within this class is best promoted by not having a patent system.

IV. Transition

Any practical suggestion to change the patent system cannot be examined entirely in absence of the transition costs of moving to such a system. Whatever the superiority of the alternative system defined above, the merits of the system must be offset by the cost of achieving such a system.

One of the potential obstacles to transition is the degree to which uniformity is required of international patent rights. The transition of an individual country away from abstract patents would not be to the benefit of the country: such a transition would put the country at a potential disadvantage relative to other countries. The benefits come from every other country raising the patentability bar, not from the individual country in raising the bar alone. Fostering international agreement to such an amendment would require some strategy, but would not be impossible: an agreement that could only be accepted unanimously ought to be sufficient to eliminate the individual hold-up problem: the country’s agreement, resulting in a beneficial patent change elsewhere, would strictly dominate the country’s hold-up, which would result in no patent change elsewhere. This consideration is somewhat naïvely simplified, however: in reality one might expect opportunistic hold-up by countries with relatively smaller gains in order to extract some of the potential surplus of countries with larger gains. So long as countries

²⁴ It might actually have a small positive effect, if the result is a patent office that operates more efficiently with fewer errors.

²⁵ Such patents show signs of being overturned even under the existing patent regime: for example, the recent *Association for Molecular Pathology v. USPTO* ruling, invalidating broad patent claims for genetic testing procedures.

could credibly commit to not giving in to such hold-up, it will still be in the interest of smaller gain countries to accept the agreement.

If firms can operate across borders with relative ease—as should be the case for software firms, for example—international agreement becomes even less important: a country's patent regime would affect sales in the country but not the decision to develop in the country. Suppose, to construct a simple example, two countries, α and β , where α allows abstract patents, while β does not. The decision by marginal firms considering selling in α will be dissuaded by the patent system disincentives, but will sell in β where the patent system does not impede sales. β will not necessarily be more likely to attract firms to develop within β : firms can still develop in α , and then sell in β without selling in α ; thus the decision of where to develop is largely orthogonal to which markets to sell in. There is, of course, a marginal benefit to closeness (in location, currency, culture, etc.) to the buyers of the firm's good, but these are again orthogonal to decision of country in which to sell: patent laws are going to primarily affect the decision to *sell* into α or β , not the decision to *develop* in α or β . Thus, whether or not α allows abstract patents, β 's optimal strategy is to disallow abstract patents, thereby increasing the available goods in β . It is worth noting that this is in stark contrast to copyright, where the Berne Convention (and later agreements) enforces international recognition of copyrights; the patent system, on the other hand, has no international recognition of foreign patents: patent clearance and infringement is judged within each jurisdiction in which the firm decides to operate.

There are two primary approaches to transition: rescinding existing patents—possibly with compensation—that would become ineligible under the new system, or by allowing natural expiration of the newly-ineligible patents. The first approach has the enormous difficulty of attempting to fairly determine appropriate compensation. At any reasonable standardized level of compensation, some patent holders will object: in particular, the successful patent trolls, the firms engaging in cross-licensing with royalties, and for firms whose patents have resulting in higher returns than had been expected when filing for the patent²⁶. Patents (and super-patents) being used defensively would be least likely to object, since the class of patent they are defending against would be invalidated along with the defensive patent.

The alternative to immediate rescinding of patents is to let patents expire at their original expiration dates. This would, however, have two effects: first, it would induce a patent filing rush as firms seek to patent under the old rules. This could be mitigated, in theory, by backdating the applicable enactment date, but such an

²⁶ There are, of course, other groups outside the innovative industries who might object: most notably, intellectual property lawyers, who would lose out in a better-defined patent environment: on the one hand, fewer patents would be sought (resulting in less attorney fees associated with the patent filing), and many fewer lawsuits alleging infringement would be filed, since the class of patents with the most uncertain infringement details—and thus highest likelihood of lawsuits—would no longer be patentable.

approach would likely simply lead to the rescinding problem again for the subset of patents filed between the selected backdate and the enactment date of the legislation. Given the typical time required for broad-reaching, international intellectual property rights agreements and implementation of those agreements, such a backdating approach would not avoid the rescinding problem at all: the period of time for patent rescinding would encompass a significant portion of the validity period for those patents. Expiry without backdating, then, would result in a rush of patents combined with a slow period of patent transition—the total transition would take twenty years from enactment. Over this period the disincentives to innovation from abstract patents would be repealed slowly, with a larger shift near the end of the twenty years depending on the size of the patent filing rush before enactment. The net result of such a gradual transition is hard to calculate: if the influx of trivial patents before the expiration date was large, the eventual increases to innovation would be offset by the increased disincentive to innovation during the transition period due to the trivial patent rush.

The optimal policy-maker decision, however, might be the rather bold move of simply ignoring the complaints by the various groups identified above and simply rescinding newly ineligible existing patents with no remuneration at all. The ones most harmed by such an action would be the patent trolls and those with unexpectedly high profits; in the former case, the policy-maker should not care: the patent troll's primary "innovation" was in creative patent claim interpretation. Eliminating this rent-seeking behaviour removes disincentives from firms targeted by patent trolls, while not reducing innovation from patent trolls, who weren't innovating anyway. In the second case—firms earning higher than expected returns—the firms are likely to have earned more than they needed to innovate in the first place, especially given the substantial period of time it would take to enact such a policy change. One potential criticism of this approach is the idea that it would unfairly treat patentees and weaken the patent system. However, since this would be a one-time change aimed at fostering increased innovation, it is unlikely that the change would detract from the patent system's perceived strength—and should, in fact, increase it. Whether such a course of action is politically feasible is a different matter: a small number of firms—the high-return tail of the distribution—would undoubtedly become worse off; a policy-maker would have to stand firm to the inevitable rent-seeking behaviour.

V. Conclusion

The rise of increasingly abstract patents, particularly due to the abstract nature of some relatively new classes of innovation, has led to a patent system that, rather than encouraging innovation, has become a drag on innovation. Taking no action—in effect choosing to maintain the current patentability of abstract ideas—discourages innovation for the industries operating among such patents. The bar of patentability needs to be revised and reset at a level at which patenting delivers

positive expected returns—thus encouraging innovation—while excluding innovation classes where patenting would discourage innovation. While such a bar already exists, it is currently set too low and needs to be increased.

The discouragement to innovation in the current patent system takes several forms. Weaker patent requirements permit a larger concentration of patents over the same innovation space, thus increasing the transaction costs in negotiation patent clearance rights²⁷. Patent systems allowing unclear or abstract patent claims introduce significant uncertainty, similarly increasing the cost of clearing patent rights, while also increasing the expected chance and cost of litigation connected with the patent. Where the patent system has failed is primarily in its failure to adapt to new classes of innovation, such as software and genetic sequencing research, who are hindered rather than helped by an aged patent system.

There have been various other proposals of potential fixes for the patent problem other than that proposed by this paper, some more feasible than others. For example, strengthening the patent office, primarily in terms of funding, thereby decreasing its error rate; making newly-issued patents easily challengeable by competitors, who might be better informed regarding prior art or other reasons a patent should not have been granted²⁸; and giving the patent office a binding “clearing” role where the patent office can make a legally-binding declaration of whether a new innovation infringes existing patents²⁹. Another, extra-patent, solutions involve a transition to a prize system for innovation. Many of these, and potentially other approaches that I mention in the following *Future research* section, have compelling justifications, and most would complement the reformation ideas suggested in this paper.

The one certainty of the patent system is that it is far from perfect. While evidence suggests that it is working well for some sectors, evidence also suggests that, for some sectors—especially younger industries dealing with more abstract, non-rivalrous products—the current patent system, through costs arising from its uncertain and unpredictable patent treatment, inhibits rather than encourages innovation. There exists no simple fix, and it may be the case that, for any reasonable discounting of future values, the present transition costs eclipse the longer-term benefits of alternative innovation-encouraging mechanisms. Nevertheless, current patent policy seems ill-equipped to provide proper incentives for burgeoning, nascent industries that deal heavily in abstract ideas rather than physical goods.

²⁷ As such, the attempt by some to associate higher patenting rates with higher innovation is particular misguided: more patents may, in fact, mean *less* innovation, if, rather than being due to larger innovation rates, the increase patenting rate is simply indicative of a miasma of patents through which navigation is costly.

²⁸ See, for example, Gallini (2002).

²⁹ Bessen and Meurer (2008).

VI. Future research

The idea of intellectual property has critical importance to the future of economic growth, and, while patents (and, to a lesser extent, other areas of intellectual property) have received substantial coverage in academic literature, the scope has often been insufficient for the explanation of encouraging the creation of intangible, non-rivalrous goods—particularly innovation such as software, which is inherently based upon abstract algorithms. As economic activity becomes increasingly governed by intangible rather than tangible creation, clearer and separate legal infrastructure is needed to encourage its continued development.

One area of future research is collecting data with which the problem can be quantitatively analysed. While some preliminary data exists, we are still lacking in data covering firms' incentives and disincentives for innovation resulting from their own patents and other firms' patents across several diverse industries. Data exists in various anecdotal instances, but the amount of consistent, uniform data available for determining firms' desire (or lack of desire) for patents would be immensely useful.

With more available data, the most obvious continuation of research would be to more closely define an appropriate dividing line between patentability and non-patentability. Indeed drawing the line between “abstract” and “non-abstract” for patents is itself a difficult task: how do we construct a rule that clearly—both for the patent office and potential patentees—delineates between acceptable and unacceptable patent boundaries? Even after answering that question, where should the line be positioned on the specific/abstract spectrum? While I have argued that the current patent system rules put the separation too far in the abstract direction, too far in the non-abstract direction could result in a patent system where patented innovations are so narrowly covered that only minor variations are required to avoid patent infringement. If, as I have argued, the bar for patentability with respect to abstractness is indeed too low, precisely how high should it be raised to alleviate the problem?

Another continuation of research in this field would be investigating alternative approaches for supporting abstract innovation outside of the patent system. Where I have identified the patent system being worse than nothing for these types of innovation, it is likely that there is a workable alternative—as already exists in the form of plant breeders' rights, industrial design rights, and circuit topography rights—that works better than having no special coverage at all. One possible, if far-fetched, approach worth consideration is borrowing a concept from copyright by introducing clearly defined “fair use” mechanism for patents, tailored to address the negative incentive problem without necessarily eliminating an entire class of innovation from patentability.

References and Further Reading

- Bessen, James and Robert M. Hunt (2004). An Empirical Look at Software Patents. Federal Reserve Bank of Philadelphia, Working Paper No. 03-17.
- Bessen, James and Michael J. Meurer (2008). *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk*. Princeton University Press.
- Denicolò, Vincenzo (1996). Patent Races and Optimal Patent Breadth and Length. *The Journal of Industrial Economics*, 44(3), 249-265.
- Gallini, Nancy (2002). The Economics of Patents: Lessons from Recent U.S. Patent Reform. *The Journal of Economic Perspectives*, 16(2), 131-154.
- Gallini, Nancy and Suzanna Scotchmer (2002). Intellectual Property: When Is It the Best Incentive System?. *Innovation Policy and the Economy*, 2, 51-77.
- Gilbert, Richard (2009). The Essentiality Test for Patent Pools. *Competition Policy Center, Institute of Business and Economic Research, UC Berkeley*.
- Hall, Bronwyn H. and Rosemarie Ziedonis (2001). The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry. *The RAND Journal of Economics*, 32(1), 101-128.
- Heller, Michael A. (1998). The Tragedy of the Anticommons: Property in the Transition from Marx to Markets. *Harvard Law Review*, 111(3), 621-688.
- Heller, Michael A. and Rebecca S. Eisenberg (1998). Can Patents Deter Innovation? The Anticommons in Biomedical Research. *Science, New Series*, 280(5364), 698-701.
- Hunt, Robert M. (2006). When Do More Patents Reduce R&D? *The American Economic Review*, 96(2), 87-91.
- Kamien, Morton I. and Yair Tauman (1986). Fees versus Royalties and the Private Value of a Patent. *The Quarterly Journal of Economics*, 101(3), 471-491.
- Katz, Michael L. and Carl Shapiro (1986). How to License Intangible Property. *The Quarterly Journal of Economics*, 101(3), 567-589.
- Levin, Richard C. (1986). A New Look at the Patent System. *The American Economic Review*, 76(2), 199-202.
- Parchomovsky, Gideon and R. Polk Wagner (2005). Patent Portfolios. *University of Pennsylvania Law Review*, 154(1).
- Scherer, F. M. (1993). Pricing, Profits, and Technological Progress in the Pharmaceutical Industry. *The Journal of Economic Perspectives*, 7(3), 97-115.
- Scherer, F. M. (2001). "The Innovation Lottery", in Rochelle Dreyfuss, Diane Zimmerman and Harry First (eds.), *Expanding the Boundaries of Intellectual Property*, Oxford University Press, 3-21.
- Scotchmer, Suzanne. *Innovation and Incentives*. The MIT Press.
- Scotchmer, Suzanne and Jerry Green. Novelty and Disclosure in Patent Law. *The RAND Journal of Economics*, 21(1), 131-146
- Sen, Debapriya and Yair Tauman (2007). General licensing schemes for a cost-reducing innovation. *Games and Economic Behavior*, 154(1), 164-186.