Economics and the Design of Patent Systems

Robert M. Hunt
Federal Reserve Bank of Philadelphia

Follow this and additional works at: http://repository.law.umich.edu/mttlr
Part of the Business Organizations Law Commons, and the Intellectual Property Law Commons

Recommended Citation
Available at: http://repository.law.umich.edu/mttlr/vol13/iss2/6

This Symposium Article is brought to you for free and open access by the Journals at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Telecommunications and Technology Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.
ECONOMICS AND THE DESIGN OF PATENT SYSTEMS

Robert M. Hunt*


I use intuition derived from several of my research papers to make three points. First, in the absence of a common law balancing test, application of uniform patentability criteria favors some industries over others. Policymakers must decide the optimal tradeoff across industries. Second, if patent rights are not closely related to the underlying inventions, more patenting may reduce R&D in industries that are both R&D and patent intensive. Third, the U.S. private innovation system has become far more decentralized than it was a generation ago. It is reasonable to inquire whether a patent system that worked well in an era of more centralized innovation functions as well for the more decentralized environment of today.

INTRODUCTION ......................................................................................457
I. PATENT STANDARDS AS INDUSTRIAL POLICY.........................459
II. WHAT IF INVENTIONS AND PATENTS ARE DISTINCT OBJECTS?..461
III. WHO IS DOING ALL THIS R&D ANYWAY? ...............................464
CONCLUSIONS...............................................................................468
APPENDIX .........................................................................................469

INTRODUCTION

Economists resort to theory as a way of sorting out the effects of different phenomena on observed economic activity. In the field of patents

---

* Research Department, Federal Reserve Bank of Philadelphia. 10 Independence Mall, Philadelphia, PA 19106. E-mail: bob.hunt@phil.frb.org.

This paper was written for the University of Michigan Law School's September 2006 conference on Patents and Diversity in Innovation. The ideas presented here have benefited from discussions with and the research of Robert Barr, John Barton, James Bessen, Wesley Cohen, Dominique Guellec, Bronwyn Hall, Brian Kahin, Mark Lemley, Robert Merges, Leonard Nakamura, Cecil Quillen, Arti Rai, Suzanne Scotchmer, and Rosemarie Ziedonis. Annette Fratantaro provided excellent research assistance. The opinions expressed, as well as any errors, are my own, and not those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.
and innovation, we observe firms making substantial investments in research and development, and we have in place very complicated, and sometimes costly, systems of intellectual property rights to encourage such investment. These patterns and institutions are so complicated that it is difficult to identify cause and effect, and so we use economic models as a way of seeing these things more clearly.

One hopes to use the right models, just like the right pair of eyeglasses, to improve one’s vision and not impair it. Of course, there is plenty of room to argue about the right way to model this phenomenon, and there is no shortage of economists to take up that space. I will set forth, in words only, a few such models and describe their implications and leave it to the reader to decide whether these clarify or muddy the picture.

I wish to make three main points. First, a uniform set of patentability standards will not be optimal for most industries in the sense that different standards might well lead to more innovation in some industries and less in others. In that case, policymakers face a tradeoff. Patentability standards should be set to maximize innovations in those industries that generate the greatest gains for the economy. If it is optimal to maximize the rate of innovation in industries where R&D is the most productive, patentability standards should be set relatively high.

Second, patents are sometimes more than the literal right to exclude others from producing the underlying invention. Once we recognize that there can be some slippage between inventions and the property rights granted by the government, it is possible that more patents can result in less R&D, particularly in industries that are both R&D and patent intensive. To mitigate such a perverse outcome, patent procedures should be refined to tighten the relationship between patents and the underlying inventions, or patent claims should be narrowed. If these are not possible, the cost of obtaining patents, pecuniary or otherwise, should be increased.

Third, over the last quarter century the private U.S. innovation system has become far more decentralized, for reasons that may have little to do with intellectual property rights. In that sense, we have more diversity in R&D than in the past, but we also have much more fragmented property rights. In addition, increased turnover within industries and more asymmetry between firms may complicate the ability of private agents to contract efficiently for access to state-of-the-art technologies. Is a patent system that functioned well in an era with a more concentrated innovation system equally effective for conditions as they are today? This is, I think, an open question for researchers.
I. PATENT STANDARDS AS INDUSTRIAL POLICY

Most economic models of patents suggest there is an "optimal" instrument in the sense that it balances two or more tradeoffs. The most typical tradeoff is between the need to provide positive profits to an innovator to recover the cost of R&D and the distortions associated with pricing the good or service above the marginal social cost of production. We can write down more complicated versions of the story, but as long as extending ever larger benefits to the innovator eventually results in diminishing returns in terms of inventive output, the story is essentially unchanged.

The design of the "optimal" patent, then, will depend on a number of parameters that describe consumers' demand for the new goods and services, characteristics of the R&D technology, market structure, and other factors. The design might be characterized in terms of patent length, the breadth of the claims, the ability to claim infringement by modest incremental improvements (what economists call leading breadth), or even the standard that determines which inventions can be patented (for example, non-obviousness, or what I will call the inventive step). The crucial point is that if industries can be characterized in terms of differences in these parameters, the optimal system of intellectual property will vary systematically across industries.

In my 2004 article, I consider an environment in which firms engage in R&D competition over many periods—in fact, forever. Thus, firms engage in cumulative innovation. Assume that patents are essential—that is, without protection, innovations are imitated at no cost at all. To simplify things, we can characterize intellectual property simply in terms of the minimum inventive step required to qualify for protection. To be concrete, innovations come in all sizes, and the size of the inventive step determines which inventions will be protected and which will fall immediately into the public domain. The higher the inventive step needed to qualify for protection, the smaller the proportion of inventions that will qualify for protection. Suppose that a patentable invention earns rents only until the next patentable invention comes along. Thus, the economic life of a patent depends on the rate of innovation in an industry

and the standard of patentability. Finally, suppose there is free entry into R&D competition. In other words, if firms anticipate making enough money, they will sink resources into an R&D facility and compete for the expected rewards.

This model teaches three lessons—one obvious and two less obvious. The first lesson is that in industries where R&D is more productive, there is more innovation and more firms willing to enter into R&D competition. Second, for any given industry described by a number of characteristics, including the productivity of R&D, there is a unique inventive step that maximizes the rate of innovation by maximizing the entry of firms. In this context, this standard is an “optimal” patent. Third, the more productive an industry is in terms of its R&D, the higher is the optimal minimum inventive step needed to qualify for protection. Thus, in order to maximize the rate of innovation in industries where R&D is highly productive, a relatively small share of innovations should qualify for protection.

The key insight from the model is that the rate of cumulative innovation depends on the degree of business’ stealing across different generations of innovation. This margin is affected by patent standards. Granting patents for too many innovations can dissipate the economic value conveyed by them, reducing the incentive to innovate. This is especially pronounced in industries where R&D is more productive because the profits dissipated would have been earned in the very near future rather than far into the future.

Now suppose we can specify only a single inventive step for the entire economy. If we characterize industries by differences in the productivity of their R&D, it is clear that any given standard will be closer to the optimal standard for some industries than for others. As a result, there will be more entry into those industries, and also more entry relative to other industries.

It is in exactly this sense that I argue that a common patent standard applied to all industries is a form of dynamic industrial policy. If one wishes to favor industries where R&D is not very productive, the inventive step should be set relatively low. The price of this policy is much

4. This is a simple way to abstract from issues of infringement and licensing. See Ted O'Donoghue, A Patentability Requirement for Sequential Innovation, 29 RAND J. ECON. 654 (1998) (considering both factors in a similar model. From his analysis, it is clear that incorporating these factors simply reinforces the intuition presented here).

less innovation in industries where R&D is more productive. This is unlikely to be a good policy for an economy whose competitiveness is based on rapid innovation. On the other hand, if one wishes to favor industries where R&D is most productive, the inventive step should be set relatively high. Even though a smaller share of inventions would qualify for protection, innovation in those industries would be more rapid, likely resulting in an increased rate of growth.

An alternative interpretation to consider is that the optimal inventive step involves balancing two competing factors: the instantaneous benefit of patenting the marginal innovation versus the dynamic losses that result from increased business stealing. In that case, one can think in terms of applying this balancing of factors to every industry, implicitly allowing for a different inventive step in every industry. This might be interpreted as a common law standard of patentability. Under this interpretation, the role of the courts is much different than under the previous interpretation.

II. WHAT IF INVENTIONS AND PATENTS ARE DISTINCT OBJECTS?

Most economic models of the patent system assume that a patent is a monopoly on the right to produce the thing that was invented. It is thus customary to think of inventions and patents as complimentary inputs in the production of profits for the firm. In the language of economists, this means that a decrease in the cost of either input would result in an increase in both inventing and patenting. This, in turn, would increase firm profits.

In my 2006 article I consider the possibility that inventing and patenting need not be as closely related as typically assumed in the literature. I then ask a simple question: “When is it the case that inventions and patents can be substitute inputs in the production of firm profits?” In other words, is it possible that a decline in the cost of obtaining patents, which would clearly induce more patenting, could result in a decline in the invention rate? The answer is yes, under a specific set of conditions made clear by the model. When those conditions are satisfied, firms will devote too many resources to patenting in order to “tax” each other’s inventions and too few resources to the process of inventing itself.

Three key ingredients are necessary to trigger this outcome. First, firms must do a sufficient amount of R&D. In other words, the perverse

outcome does not explain why some firms and industries do not do R&D. But it may explain why a fairly R&D intensive industry is doing less R&D than one might expect, given its productivity and other fundamentals, or relative to the past. Second, patents must be cheap relative to both the cost of doing R&D and in terms of the value of final output in the industry. Finally, there must be sufficient technological overlap between firms so that there is a good chance that one firm’s invention may infringe a patent obtained by another.

The first two conditions describe, respectively, the characteristics of an industry and the resource cost of obtaining patents. The third condition requires some additional explanation and interpretation. For some industries the third condition may simply be a question of technology: firms might draw from similar technical fields and arrive at similar solutions even when they apply them to different problems. This is particularly true for industries that advance through cumulative innovation and where firms rely on a largely common set of building blocks derived from previous innovations. In addition, some products incorporate several, if not dozens, of potentially patentable innovations. Two obvious examples of this type of industry are the semiconductor and computer software industries.

The degree of overlap might also depend on the breadth of the patent claims themselves. If broad claims are regularly granted, it is more likely that firms will infringe each other’s patents. Under this interpretation, it is possible that the patent office, the courts, or, in the last instance, the legislature could regulate the degree of overlap.

A third, and more controversial, interpretation is that firms have some facility in obtaining property rights over things they have not yet invented. While this is prohibited by U.S. patent law, it might nevertheless arise from mistakes in the examination of patent applications. This is a topic that has received considerable attention in academic and policy circles in recent years.

This may be a particular problem for patents on computer programs, especially ones that implement methods of doing business, if patent

---


law's disclosure requirements are not adequately enforced. In that case, one might not be certain what the applicant has invented and how far his or her claims should extend. In these areas, some researchers and practitioners worry that applicants can obtain relatively broad patents even though they have not really started their R&D.

The model suggests that the perverse outcome is more likely to occur in high-tech industries that advance rapidly via cumulative innovation (and which also obtain many patents) than in industries that are not research intensive or which do not build up large patent portfolios. Previous empirical work has identified a number of industries with such characteristics, including the electronics, computer, and semiconductor industries. These industries account for most of the rapid growth in U.S. patenting in recent years. They are also the industries where researchers identify what is sometimes called "strategic patent" behavior, including the assembly of large portfolios of patents for wholesale cross-licensing and, in some instances, deceptive patent prosecution.

In our 2004 working paper, Jim Bessen and I present empirical results in the context of patenting computer software that are consistent

---

9. For disclosure requirements, see 35 U.S.C. § 112 (2000) (patent applications must "contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention").


with the phenomenon modeled here. Obtaining such patents was difficult, but not impossible, during the 1970s and early 1980s. Over time, however, courts became more receptive to such patents, and their numbers have grown rapidly, especially among firms in the industries described above (and much less so among firms in the software industry). All else equal, firms that concentrated on obtaining software patents experienced a statistically and economically significant decline in their R&D intensity relative to other firms.

If there are instances where patents substitute for R&D, what is the appropriate policy response? It is clearly most important to modify the patent process to ensure that there is a closer relationship between what a firm invents and the property rights it subsequently obtains from the government. As described earlier, this may involve modifications to patent law’s disclosure and enablement requirements. It might also involve new re-examination procedures or some other process of objecting to recent patent grants. It could also involve a modification to the presumption of patent validity accorded in trials so that courts could engage in a more reasonable assessment of the quality of a patent’s examination.

If the breadth of claims is the issue, granting more narrow patents would likely help even if this resulted in more patents being granted. But, as noted earlier, overlapping claims may be an unavoidable outcome for some technologies or for some industries whether innovation is cumulative in nature.

Finally, if these other avenues are not effective, another option is to raise the cost of obtaining patents. This might involve higher pecuniary charges, but it might also involve raising the standards used to determine when an invention is eligible for patent protection. In other words, a higher inventive step might also mitigate the problem.

III. WHO IS DOING ALL THIS R&D ANYWAY?

I have already described a relationship between patent standards, the number of firms engaged in R&D, and the overall pace of innovation. In more recent work, Leonard Nakamura and I return to the question of

---


15. This result is reported in the working paper version of my theoretical model. Robert M. Hunt, When do More Patents Reduce R&D (Fed. Reserve Bank of Phila., Working Paper No. 06-6, 2006).
exactly who is doing R&D and why. We document the very significant change in the distribution of private R&D investments across the U.S. economy and attempt to sort through a number of competing explanations for these changes. One conclusion from this work is that the U.S. innovation system is vastly more decentralized than in the past. We believe the increased decentralization is largely unrelated to the patent system. An important question, then, is whether our current system of intellectual property rights and the methods for enforcing them are well suited to this new decentralized innovation system.

The U.S. economy can be divided into industries that are R&D intensive (R&D is at least one percent of sales) and those that are not. If this definition is applied to firms in 1973 (when accounting rules required the disclosure of material amounts of R&D), does the same division of industries remain accurate twenty-five years later? The answer, somewhat surprisingly, is yes (see Figure 1). Nearly all private R&D continues to occur in the same R&D-intensive industries we observed back in 1973. If anything, R&D is more concentrated in these industries, since the R&D intensity of these industries has risen significantly over time (see Figure 2).

One can also examine the distribution of R&D within an industry. This also changed dramatically after 1980. We examined a particular set of large, established firms, about 65 in number, which accounted for the majority of all private R&D around 1973; we call these firms the incumbents. As Figure 3 shows, these firms accounted for about the same share of private R&D until about 1980; thereafter, their share fell to about thirty-five percent of the total. A similar analysis, based on the size distributions of public and private firms in the National Science Foundation survey of industrial R&D, tells the same story (see Figure 4). Thus, the decline of particular firms in our list of incumbents does not adequately explain the trend.

Large, established firms are not doing less R&D than in the past. Indeed, on average, they are doing more R&D, and their R&D intensity has grown over time. But the R&D intensity of smaller, younger firms—which we call the entrants—has grown even more dramatically (see Figure 2). Indeed, it is the rise in R&D intensity among this group of...

17. Tables 1 and 2 are derived from Standard & Poor's Compustat. We report R&D intensity as R&D divided by the operating expenses of firms because we do not wish to exclude from these calculations firms engaged in extensive R&D but not yet producing significant revenues.
18. The differences in the height among the first set of bars is statistically significant—in other words it is clear that incumbent firms used to be more R&D intensive than other firms. The differences in the height among the last set of bars is not statistically significant—
firms that explains nearly the entire rise in the R&D intensity of the U.S. economy after 1980 (this is also evident from Figure 4).

What explains these patterns? In the Article, we develop a simple model that permits us to sort among a number of competing explanations by examining differences in their predictions about trends in the observable data. Setting aside all the details, the crucial point is that the patterns in the data are best explained by a decline in the fixed costs firms incur after the success of their R&D projects. These are the costs firms pay to reach a national or international market for their goods and services. In the Article, we focus on the decline in the costs of administration, distribution, communication, and other costs of coordination that resulted as computing power became both scalable and ubiquitous. We describe these as investments in marketing capital. A decline in the cost of marketing capital leads to an increase in R&D competition among both smaller and large, established firms. All else equal, this increased competition reduced the market value of the incumbent firms and increased the value of the entrant firms.

Why is this digression relevant to patent policy? First, the trends we observe in R&D and other observable measures could be explained by dynamics other than patent policy. For example, in some circles the overall rise in R&D intensity of the U.S. economy is taken as evidence of the success of Congress and the Court of Appeals for the Federal Circuit in improving our patent system. But without controlling for these other dynamics, it is extremely difficult to establish a casual argument about the effect of the earlier changes in patent policy or aggregate measures of investment in R&D.

The second point is that the private U.S. innovation system has undergone a dramatic structural change. It is unclear whether our patent system is as effective in this new environment as it was forty years ago, when R&D activity was more concentrated. There are also more independent research programs simultaneously pursuing solutions to similar technical problems. So in that sense, we have more diversity in R&D programs than in the past. But it is almost certainly true that some of these programs, and thus their results, overlap.

Now one could argue the democratization of R&D is an outgrowth of the "strengthening" of intellectual property rights in the U.S. after 1982. The argument is that increased certainty about enforcement and enhanced penalties for infringement facilitates entry because firms are better able to protect their intangible investments, obtain financing, and in other words we cannot be sure there are significant differences in the R&D intensity of these groups of firms. Nevertheless, we can conclude that younger and smaller firms have at least caught up with the incumbent firms in terms of the research intensity.
license their technologies.\textsuperscript{19} This is an appealing argument, but one for which there is relatively little empirical evidence.\textsuperscript{20} But even if this is an important channel, it is complementary to our story since declines in the cost of marketing capital will improve the terms of trade for young firms in their licensing negotiations with established firms. Nor is there a consensus among economists that these changes in the U.S. patent system have favored younger, smaller firms over larger, more established firms.

Regardless of how it came about, the decentralization of R&D poses a number of potential challenges to our patent system. What is the best way to coordinate a patchwork of property rights that are owned by a much more numerous and diverse population of inventors? Can we rely on Coasian bargaining to resolve these conflicts without additional government intervention? One reason to think not is that, due to the combination of a more rapid turnover of firms and an increased asymmetry between firms, the mutual forbearance we expect from repeated interactions within an industry may be undermined. In such environments, activities such as licensing, patent pools, and standard setting all become more important.\textsuperscript{21} While there is an extensive theoretical literature that studies these activities, the existing empirical literature is rather thin.

I will not summarize the conclusions from this literature except to make two points. First, one should not assume the associated transaction costs are minimal or, for that matter, impossibly high. Transaction costs will undoubtedly matter, and it is up to us to find out how large they actually are.\textsuperscript{22} The second point is that the effects of each of these mechanisms (licensing, patent pools, and standard setting organizations) depend on the quality of the property rights being granted. Facilitating the licensing or pooling of bad patents is probably bad policy.\textsuperscript{23} A largely

\begin{footnotesize}

\begin{enumerate}
\item \textsuperscript{19} I thank Wesley Cohen for making this point. Our empirical analysis attempts to control for these effects by, among other things, taking into account firms' patenting activity over time.
\item \textsuperscript{20} One exception is the work of Hall and Ziedonis on the semiconductor industry. See Hall & Ziedonis, supra note 11. See also Ashish Arora et al., Markets for Technology: The Economics of Innovation and Corporate Strategy (2001).
\item \textsuperscript{21} In addition, firms may respond to fragmented property rights by accumulating even more patents—bargaining chips for use in subsequent licensing disputes. See generally Rosemarie H. Ziedonis, Don't Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms, 50 MGMT. Sci. 804 (2004).
\item \textsuperscript{22} For a discussion, see Fed. Trade Comm'n, supra note 8.
\end{enumerate}

\end{footnotesize}
unexplored question is whether the optimal quality of patent examinations should depend on the market structure of the private innovation system. If so, it is reasonable to consider whether our examination process should be revised to reflect the current environment.

Two additional factors may be relevant. First, what criteria should be used to determine when an injunction should issue, especially if the questions of validity and infringement are not yet decided? Second, what is the appropriate method of calculating damages for infringement in an environment where one firm has a patent and nothing else, and another has made substantial investments in order to bring a product to market under conditions of significant uncertainty? Should the rules depend on whether there are many active firms or only a few?

CONCLUSIONS

This Article reviews a number of economic analyses and offers several tentative suggestions about the design of our patent system. First, unless a balancing test reflecting variations in industry/technological characteristics can be implemented, a uniform set of patentability criteria will favor innovation in some industries over others. It is up to policymakers to determine the sectors in which innovation should be maximized—presumably where R&D is the most productive. Second, for industries that are both R&D and patent intensive, it is vital that the property rights granted conform closely to the innovations they seek to protect. This may require changes to the patent examination process, the presumption of patent validity, and the interpretation of patent claims; the implementation of a post grant opposition process; and possibly increasing the pecuniary and other costs associated with obtaining patents. Third, given that the private innovation system in the U.S. is far less centralized than it was a generation ago, more research is required to determine how efficiently the private sector can coordinate and price access to technologies developed by an extremely heterogeneous pool of inventors. While more theoretical research is needed, the primary impediment to understanding the efficacy of these mechanisms is the dearth of empirical work. This needs to change.

APPENDIX

FIGURE 1
PRIVATE R&D REMAINS IN THE "OLD" R&D INDUSTRIES*

![Bar chart showing percent of all private R&D from 1974 to 1999.](chart1)

Source: Hunt and Nakamura (2007)
*
*: 3-4 digit SIC industries with R&D/Sales ≥ 1% in 1973

FIGURE 2
PRIVATE R&D INTENSITY RISES OVER TIME

![Line chart showing R&D as a share of operating expenses from 1974 to 1999.](chart2)

*: Incumbents are firms with 25,000+ employees in 1965


**FIGURE 3**

**THE SHARE OF PRIVATE R&D FROM INCUMBENT FIRMS DECLINES**

![Bar chart showing the share of private R&D from incumbent firms declining from 1974 to 1999.](image)

Source: Hunt and Nakamura (2007)

*: Incumbents are firms with 25,000+ employees in 1965

---

**FIGURE 4**

**CONTRIBUTION TO U.S. PRIVATE R&D INTENSITY BY FIRM SIZE**

![Line chart showing the contribution to U.S. private R&D intensity by firm size from 1959 to 1999.](image)

Source: NSF Survey of Industrial R&D and author's calculations