Patent Theory versus Patent Law

Alexander Tabarrok
George Mason University
Patent Theory versus Patent Law

Abstract

According to the economic theory of patents, patents are needed so that pioneer firms have time to recoup their sunk costs of research and development. The key element in the economic theory is that pioneer firms have large, hard to recoup, sunk costs. Yet patents are not awarded on the basis of a firm’s sunk costs. Patent law, in fact, ignores costs. The disconnect between patent law and patent theory suggests either that modifying patent law so that it better fits with patent theory would reduce the costs and inefficiencies associated with current patent practice or that the standard economic theory of patents is wrong.
1 Introduction

An often repeated argument for patents is that by giving inventors a limited monopoly in their inventions the “progress of Science and useful Arts” is promoted. Stated differently, the prospect of monopoly profits increases the incentive to innovate. It’s not immediately obvious why society would want more innovation than would occur without patents. It’s quite possible to invest too many resources in research and development rather than too few.

Economic theory, however, provides an argument for why patents could improve the allocation of resources. Original research and development is usually more costly than imitation. A firm will not be able to recoup its sunk costs if the results of its research are quickly imitated by rivals. Recognizing this, firms will have little incentive to invest in innovation. Patents and other forms of intellectual property increase the incentive to innovate by delaying the arrival of imitators thus giving pioneer firms time to recoup their sunk costs through monopoly pricing.

The “recouping the sunk costs of innovation theory” is the dominant theory of patents among economists (hence, I will also refer to this as the economic theory). Thus Bessen and Maskin (1991) write, “The standard economic rationale for patents is to protect potential innovators from imitation and thereby give them the incentive to incur the costs of innovation.” Similarly, Henderson (2002) notes that:

The economic justification for patents is straightforward...If imitators have the same production costs as the inventor, they could compete the price down so that the original inventor covers only production costs, but not invention costs. Potential inventors, knowing this, would be less likely to invest in inventing.

The economic theory dates back to at least Jeremy Bentham, who wrote that “the protection against imitators” is necessary because “he who has no hope that he shall reap will not take the trouble to sow.” Arrow (1968) offers an influential modern restatement that focuses on the general non-appropriability of information. Nordhaus (1969) first formally modeled the tradeoff between innovation and monopoly distortions. (Machlup 1958 and Menell 2000 survey the older and more recent literatures respectively.)

1 Quoting the U.S. Constitution Article 1, Section 8.
Although the economic theory is well accepted, I argue that it does not fit well with the actual patent system. A patent system designed around the idea of recouping sunk costs would look quite different than the current system. In particular, as I will explain below, the current system ties returns to innovation to the benefits of a patented idea, i.e., as the value of the idea increases monopoly profits increase. Yet the economic theory of patents implies that returns should be tied to the sunk costs of researching and developing the patented idea.

The poor fit between the economic theory of patents and the actual patent system suggests two alternative hypotheses. Either the theory is correct and the patent system is poorly designed, or the patent system is well designed but not for the purposes of recouping sunk costs. Which of the two hypotheses one prefers depends in part upon one’s priors about the relative efficiency of theory versus practice. My primary approach is to assume that the economic theory is correct. Thus, I argue for reforms that would bring the patent system more into accordance with the economic theory. In section 7, however, I briefly take up the possibility that the theory may be wrong and the patent system optimal for other purposes. If one’s priors are that the patent system as it exists is efficient, even if we do not understand why, then my arguments for reform can be understood as testing the economic theory. To the extent that the patent system could be reformed so as to make it more consistent with the economic theory, the fact that we have not done so suggests a fortiori that the economic theory is wrong.3

2 Theory

Consider the following simple model. Let $B$ denote the social benefits to the intellectual property, should it come into existence, and $C$ the private costs of creating the intellectual property. I will assume that $C$ is primarily a sunk cost and that once the intellectual property has been created it can be distributed at relatively low marginal cost. Pharmaceuticals, software, microchip designs, films and musical recordings fit this paradigm to a large degree. Inventors capture only a fraction $\alpha$ of the benefits of intellectual property but they bear all the costs. Intellectual property will be produced so long as $\alpha B - C > 0$ or, rearranging, so long as $\alpha > \frac{C}{B}$.

If it were possible to set $\alpha = 1$, this would maximize the social gains from

---

3 A third hypothesis is that the economic theory is consistent with the existing patent system given that additional factors are taken into account. It is, of course, always possible to rescue any theory with enough supplementary factors.
innovative activity (ignoring distributional issues). In practice, however, increasing $\alpha$ typically requires an increase in monopoly power. Thus, an extra dollar in benefits to innovators comes at the expense of more than a dollar lost by consumers. As a result $\alpha$ is always less than 1. If the maximum feasible $\alpha$ is less than $\frac{C}{B}$ the innovation will not be produced even if $B > C$. If the maximum feasible $\alpha$ is greater than $\frac{C}{B}$ then $\alpha$ should be set for each innovation at the minimum level consistent with the project being profitable, i.e., at $\alpha = \frac{C}{B}$.

An immediate implication of the model is that $\alpha$ should be low when $B$ is high.\(^4\) Note how this conflicts with the often heard intuition that intellectual property should be highly protected when it is valuable and not so highly protected when it is of low value. Yet in an optimal system high value property is weakly protected and low value property strongly protected.

The current patent system rewards innovators with monopoly profits. If costs varied strongly with benefits, as is the case for non-information goods, then rewards would vary (indirectly) with costs. But in the case of intellectual property the relevant costs are sunk and thus do not vary with benefits. As a result, the patent system often offers innovators large rewards despite the fact that research and developments costs are small. That is, even though the economic rationale for patents is to allow innovators to recoup R&D costs it often occurs that $\alpha B$ far exceeds $C$. Absent the rationale of recouping R&D costs, intellectual property protection creates monopolies with attendant dead weight losses and no social benefits.

The granting of patents when $\alpha B$ is far greater than $C$ is not due to the difficulty of comparing $aB$ with $C$, which would sometimes occur even if the system were designed on optimal principles. Rather the problem is that the

\(^4\)The model in the text is probably the simplest one that captures the standard theory of patents. The standard theory, however, lives alongside other theories that are seemingly similar but actually quite different. An alternative theory, for example, holds that patents are necessary in order to speed the process of creation, or to say much the same thing to increase probability of creating an innovation per unit of time. In a war, for example, the government might want to offer strong patent rights (high $\alpha$) for war-related innovations even if fixed costs were low or imitation was not likely. Putting aside the extraordinary circumstances of war, however, it’s clear that faster is not always better. At some point, diminishing returns set in and the costs of increased speed exceed the benefits. Justifying a “speed theory of patents” requires, therefore, a theory of externalities to explain why the ordinary incentives provided by the market are not adequate.

I do not take up alternative theories here because in general they would not contradict the model in the text, although they would add other factors that would have to be taken into account in any application. The model in the text can be thought of as a partial model that focuses attention on certain considerations in the theory of patents that are not usually discussed.
patent system pays almost no attention to \(^C\). U.S. patent law, for example, requires a patentable innovation to be novel, useful and non-obvious\(^5\) but there is no requirement that profits and costs be compared or even that the patentee have incurred substantial sunk costs. \textit{Thus, U.S. patent law has few connections to the economic theory of patents.}\(^6\)

If U.S. patent law were based on the economic theory of patents we would expect it to take into account sunk costs (more precisely, the costs that innovators face that imitators do not). Yet many thousands of patents, have been granted where the sunk costs of development are low or even nil. Such patents have become especially common since the courts expanded patent protection to software, business and medical process patents.

3 Practice: Examples of Patented Products with Low Innovation to Imitation Costs

Amazon’s “one-click purchasing” patent (no. 5,960,411) was granted to Amazon for computer software that stores a customer’s address and credit card number in a database and then allows the customer to order items with one-click. The patent is controversial and has been challenged on the grounds that it is obvious and that prior-art existed. From an economic point of view, however, these legal arguments miss the main defect of the patent. The one-click technology does not involve extensive R&D costs. The writing of software does involve some sunk costs (these are relatively small in this case) but these costs are faced by pioneers as well as by imitators. Amazon’s innovation of one-click purchasing does not substantially lower the cost to Barnes and Noble of creating similar software. As there are few sunk R&D costs to recoup and no free rider problem, these sorts of innovations should not be patentable under the economic theory of patents. Yet, if Amazon’s claim holds up they have been granted a twenty year advantage over their rivals at the expense of consumers everywhere.

Patents on medical procedures were illegal until 1954 and rare until re-


\(^6\) There is one minor exception. A handful of court rulings on patent disputes have used expenditures on research and development as an indicator of “non-obviousness.” See Merges (1992) who cites among other cases, Panduit Corp. v. Dennison Mfg. 774 F. 2d 1082, 1099 (Fed. Circ. 1985) (fact that patent holder spent years of time and millions of dollars on research is evidence that prior art did not render invention obvious)- vacated on other grounds 475 U.S. 809 (1986); and Hardinge Bros v. Marr Oil Heat Mach. Corp., 27 F. 2d 779, 781 (long and expensive investments evidence that invention was not obvious.)
cent times but today they are becoming common. Patent number 5,080,111, a description of a self-sealing incision for use in eye-surgery was granted to Samuel Pallin in 1992. Pallin then sued other doctors to obtain royalties for use of his procedure. By Pallin’s own account he “invented” the procedure when he made an incision in a patient’s eyeball but didn’t have time to stitch the cut because the patient experienced heart problems and had to be rushed to hospital (Shulman 1999, 38). When the eye was examined several weeks later he discovered that the scar had healed spontaneously. Pallin then sought and received a patent on the type of incision that he had used. Essentially no R&D costs were incurred in Pallin’s invention, which is better described as a discovery. Yet lack of R&D is not a bar to patenting.

It is not surprising that medical procedures can be patented on the basis of U.S. patent law because a medical procedure can be just as novel, non-obvious and useful as say a medical device. The patenting of medical, business and other procedures is a logical extension of U.S. patent law. Yet the extension is unwarranted when the “invention” does not require the outlay of substantial sunk costs. Novel, non-obvious and useful ideas should be patentable only to the extent that patenting is necessary to allow inventors to recoup their sunk costs and thus to profit from their inventions.

The disjunction between patent law and patent theory has become more evident as patent law has become more liberal towards the patenting of ideas. Thomas Edison invented and patented numerous products: the lightbulb, the phonograph, movie film and much else besides. The invention of products typically requires the expenditure of sunk costs in a way that the creation of ideas does not. At one point the patent office required that patents be accompanied by working models but today it is not necessary to implement an idea to patent it and many patentable ideas are so broadly phrased that they could not be implemented in a single model. Jerome Lemelson, for example, is second only to Thomas Edison in the number of patents issued to him (Shulman 1999). Yet the public would be hard put to name a product that Lemelson has produced. Instead, Lemelson has made a fortune by patenting ideas, often long before their technical imple-

---


In part in response to the Pallin patent (see text), a bill was drafted to make patents on medical procedures illegal in 1996 (as is the case in Britain, Canada, Japan and 80 other countries) but after intense opposition from lawyers and biotechnology companies a compromise was reached that recognized medical procedure patents but made them unenforceable against doctors and hospitals (35 U.S.C. §287(c), the legislation does not extend to biotechnology patents.)
mentation into products. Lemelson, for example, patented the idea of the video camcorder but did not build any camcorders (patent no. 4,819,101). The most infamous Lemelson patent is especially illustrative of the problems created when patent law is divorced from the economic justification for patents. Patent no. 5,351,078 for “Apparatus and Methods for Automated Observation of Objects” was originally filed in 1954 but due to delay at the patent office and a series of emendations the patent was not granted until 1994. The patent gave Lemelson rights to the concept of machine vision, i.e. (quoting from the patent abstract)

“a detector, such as a camera or radiation receiver, moves around an object, which is supported to be rotatable such that the detector may receive electromagnetic energy signals from the object from a variety of angles...The detector generates analog image signals resulting from the detected radiation, and an electronic computer processes and analyzes the analog signals and generates digital codes, which may be stored or employed to control a display.”

The patent contains no technological innovations. As with the camcorder patent, the machine vision patent is little more than a coarse description of the major parts and functions of a machine vision system (e.g. “a detector of electromagnetic radiation,” “an electro-mechanical assembly controllable to cause said support and said detector to move relative to each other”, “a storage device coupled to the generator capable of storing the image signals”).

Edison famously said that “Genius is one percent inspiration, ninety-nine percent perspiration.” A patent system should reward the ninety-nine percent perspiration, not the one percent inspiration. In inventing the lightbulb, for example, Edison laboriously experimented with some 3000 possible materials for the filament, before hitting upon carbonized cotton thread (Shulman 1999). If Edison were to patent the lightbulb today he would not need to go to such lengths. Instead, Edison could patent the use of an “electrical resistor for the production of electromagnetic radiation,” a patent that would have covered oven elements as well as lightbulbs.

By the time Lemelson’s patent on machine vision was issued there were already hundreds of thousands if not millions of robot systems that used

---

8 The Columbia World of Quotations (1996) cites Edison as having made this remark circa 1903 on the authority of Harper’s (New York, Sept. 1932).
9 For a much more modest example of patenting a 1%-inspiration idea without the 99%-perspiration investment, see note 13.
electromagnetic energy to identify and help manipulate objects. Lemelson did not build these systems and the development of machine vision owes nothing to his patent, the potential existence of which was secret until 1994. Nevertheless when the patent was granted thousands of firms became instant infringers and as a result the patent has generated hundreds of millions in royalties.\textsuperscript{10}

It's quite possible for $\alpha B$ to be much larger than $C$ even when $C$ is not negligible. This can occur when the patent granted is excessively broad. Consider Agracetus’s European patent (no. 301,749) on genetically altered soybeans. Agracetus is a biotechnology firm that invested substantial resources in perfecting a method to insert foreign genes into soybean seeds to produce better phenotypic traits. A patent is a reasonable method of preventing imitators from free riding on their technique and allowing Agracetus time to recoup sunk costs. Yet the patent actually granted gave Agracetus the rights not just to the soybean they invented or to the technique they perfected but to any genetically altered soybean created by any method.

Patent law provides little guidance on how broad a patent should be. In practice, much depends on persistent lawyering and lobbying at the patent office. Economic theory, however, does suggest a standard for how broad a patent should be: holding patent duration constant, the patent should be just broad enough so that the sunk costs of innovation can be recouped through monopoly profits.\textsuperscript{11} Thus costs should be a key element in what

\textsuperscript{10}Lemelson’s patent is an example of a so-called submarine patent - a patent that surfaces years after an industry has been in operation and that automatically puts companies in infringement. Submarine patents have been unsuccessfully challenged because of delay but the real problem is that there are rarely any substantial fixed costs associated with these inventions.

What is important about submarine patents is that a patent that surfaces after the product it claims to invent is already in use cannot have encouraged the creation of that product - the very raison d'etre of patents. There can be no economic justification for patents that redistribute wealth but do not encourage wealth creation. Yet in Kingsdown Medical Consultants, Ltd. v. Hollister Inc. (863 F.2d 867, 874, 9 USPQ2d 1384, 1390 (Fed. Cir. 1988), cert. denied, 490 U.S. 1067 (1989)) the court noted that it is not “in any manner improper to amend or insert claims intended to cover a competitor’s product the applicant’s attorney has learned about during the prosecution of a patent application.”

The incentives for patentees to delay patents while amending them was reduced in 1995 when the patent law was changed so that a twenty year clock on the patent begins ticking once a patent has been filed (as opposed to the previous 17 years after acceptance rule). Even more importantly, in November of 2000 the U.S. harmonized its patent disclosure law with world practice so that patent applications are now made public 18 months after the filing date.

\textsuperscript{11}Klemperer (1990) and Gilbert and Shapiro (1990) examine a different aspect of the breadth question. These authors ask what is the optimal tradeoff between patent length
patents may claim but there is not yet a recognition in patent law for this approach (see further below).

Many of the above patents are controversial and are being challenged in court. It’s important to note, therefore, that these examples are illustrative of general patterns and could easily be expanded. (The number of software patents, for example, has increased more than 10-fold in the last decade.) If any of the above patents are invalidated it will have to be done by finding a problem specific to each patent such as the existence of “prior art.” The objections raised here, however, suggest that the ideas and inventions discussed above should not be patentable in principle.

4 Proposal

The problems described above can be alleviated by reforming the patent system so that it is brought into greater consonance with the economic theory of patents; this requires that sunk costs be incorporated in some aspect of the granting of patents. At the most basic level the issue is to adjust policy so that the benefits of the innovation captured by the innovators, $\alpha B$, are larger than the costs of innovation borne by the innovators, $C$, but not excessively so. Patent policy primarily affects $\alpha$ and $\alpha$ can be adjusted in a variety of ways including changing (a) what products can be patented, (b) the probability that a patent is granted or enforced, (c) the length of the patent or (d) the breadth of the patent. (Tax policy can reduce $C$ a point I will return to briefly below). For the purposes of incorporating sunk costs it doesn’t matter which of these methods is chosen - the critical issue is the total return to the innovators and not how that return is earned.

The most modest proposal would be to choose what sorts of products can be patented based upon an examination of typical innovation costs relative to imitation costs in that product category. Two of the most controversial new areas of patent protection, software patents and business methods patents would likely fail the sunk cost test. More generally, the economic theory of patents suggests that patent examiners and judges should be especially skeptical about patenting an idea when the genesis of the idea required few

---

and patent breadth given that total profits to the innovator are held fixed. The problem discussed here is given that patent length is fixed, at say 20 years, then how broad should a particular patent be? The two issues are different and a unified theory has not yet been developed.

12 Note that software itself is protected by copyright - software patents apply to the ideas implemented in the software.

http://www.bepress.com/bejeap/contributions/vol1/iss1/art9
expenditures. If research and development costs are not substantial then the ordinary advantages that a firm receives for providing a better product at lower cost are enough to motivate socially beneficial investments. The market norm is that producers are their costs (this is true in a competitive equilibrium). A departure from the market norm is warranted only when the costs of imitation are substantially less than those of innovation. The modest version of the proposal can thus be thought of as a minimum requirement - at a minimum no patent should be granted if research and development costs are low.

To go beyond the modest proposal requires a brief return to theory. In a world of costless information, patent authorities could assign patent duration (and other factors such as broadness) such that \( \alpha = \frac{C}{B} \), at least in expectation, for each innovation. Information is too costly and estimating procedures are too fragile, however, to expect this approach to work. But note that it’s much easier to estimate \( C \) than \( B \) because the relevant costs have already been paid at the time the duration decision must be made while \( B \) is an uncertain future benefit. If the ideal is not possible it may

---

13 It may be useful to describe the genesis of this paper. At the height of the Internet boom in early 2001, I came up with the idea of incorporating small bar code scanners within web-enabled telephones and PDAs. When shopping in a physical store these devices would let the user easily scan a product’s UPC (bar code). Internet-based software would then look the product up in a database and present the user with a list of stores and prices at which it was available. If the online price were substantially below the price in the physical store the user could on-the-spot buy the product online for home delivery.

I did not spend years or even days thinking up this idea, it occurred to me while I was taking a shower. I did spend a couple of hours confirming the existence of UPC databases and cheap bar code scanners. On a lark, I investigated whether I could patent the idea. Not having invested much effort in the idea I was only slightly disappointed to find that exactly this idea had been patented just a few months previously (patent 6,134,548). When I read the patent, however, I was shocked to find that there was little more to it than I had independently devised while showering. I had thought my idea mildly clever, albeit not nearly so clever as many of the ideas one can find written up in Popular Mechanics or displayed in science fiction novels and movies. It seemed incredible, therefore, that anyone could get a 20 year lock on any implementation of this idea simply by writing it down and submitting it to a government bureau for approval. I am aware of the danger that sour grapes may make for a sour whine.

14 The benefits of the patent, \( B \), are uncertain at the time of the innovation but they become more certain as time passes. This motivates the following idea: Register innovations at the time they are made but award no patents. Twenty years after the innovation is registered, give prizes to the innovators based upon estimates of both \( B \) and \( C \). With sufficient payment of interest, there is no cost to delay and much to be gained from the additional information that delay reveals. Abramowicz (2003) usefully surveys recent ideas.
yet be feasible to use information about $C$ to improve the patent system. If we do not charge the patent office with estimating $B$, however, then there are two areas where slippage from the ideal is unavoidable.

Recall from the theory section that, holding all else equal, innovations with low benefits should be given long patents and innovations with high benefits should be given short patents. But if the patent office assigns patent duration based on costs rather than costs and benefits then it will mistakenly assign patent duration for two sorts of innovations. With perfect information, some innovations with low innovation costs would nevertheless be assigned strong patent rights if the benefits of these innovations were so low that even low costs could be recouped only with strong patent rights. A patent system based on costs will “mistakenly” grant short-term (perhaps zero term) patents to these infra-low benefit innovations. Similarly, with perfect information some innovations with high costs would nevertheless be assigned weak patent rights if the benefits of these innovations were so high that even high costs could be recouped with weak patent rights. A patent system based upon costs only will “mistakenly” grant long-term patents to these supra-benefit innovations.

It is fortunate, however, that both of these errors are of low cost relative to the costs of the current patent system. An infra-low benefit innovation is by definition an innovation with very low benefits so the fact that some such innovations will be deterred because, under the proposal, the patent office grants them patents of short duration is a small loss. It is costly to grant supra-high innovations long-duration patents when short-duration patents would be enough to bring these innovations into being but the current system already makes this error so relative to the current system there is no loss under the proposal.\footnote{In principle, a patent system with variable duration (see below) could offer patents of duration longer than 20 years for very high-cost innovations but Nordhaus’s (1969) estimates suggest that in practice the benefits of extending duration beyond 20 years are small.} In contrast to the current system, the proposal will at least take into account one of the factors, sunk costs, that create variation in the optimal patent duration. High cost innovations will receive long patent times under the reformed system, just as with the current system, but under a reformed system low cost innovations will receive short patent durations resulting in potentially large increases in social welfare.

The U.S. patent system currently offers only two types of protection, 0 or 20 years. A more flexible system would let inventors opt to apply for a...

\footnote{In principle, a patent system with variable duration (see below) could offer patents of duration longer than 20 years for very high-cost innovations but Nordhaus’s (1969) estimates suggest that in practice the benefits of extending duration beyond 20 years are small.}

http://www.bepress.com/bejeap/contributions/vol1/iss1/art9
patent of varying length, say for 3, 8, or 20 years, with the understanding that scrutiny - based upon an examination of sunk costs - would increase as the length of the applied for patent time increased. Under such a system, patentees with low sunk costs would self-select towards patents of shorter lengths, thereby greatly reducing the deadweight losses caused by the patent system without impeding innovation.

A number of countries\textsuperscript{16} already have experience with a self-selected system of variable-duration patents, although these differ in an important way from the system that I propose. Petty patents, also known as “utility models” grant substantially similar protection as patents but for shorter periods of duration, usually 7-10 years. What distinguishes petty patents from full patents is that petty patents need not pass as strict a “non-obvious” standard as full patents. In contrast, the proposal that I have made is that the distinguishing factor be not non-obviousness but expenditure on sunk costs. Although expenditure on sunk costs can be a good indicator of non-obviousness - as the courts have recognized (see note 6) - an invention created with low innovation to imitation costs should not be patented even if non-obvious.\textsuperscript{17}

A disadvantage of patent reform in the duration dimension is that this would require new patent legislation. Patent breadth, however, is far more subject to the discretion of patent examiners and the courts. It may be easier, therefore, to bring sunk costs into play in the allowing of patent claims and in infringement decisions than in duration decisions. The basic law of patent claims is that an innovator may write the claims as broadly as he likes so long as the claims are not covered by prior art and the specification enables one skilled in the relevant art to make and use all of the embodiments of the invention encompassed by the claims. Claims, however, may substantially exceed the embodiments of the innovation that are actually disclosed in the

\textsuperscript{16}According to the World Intellectual Property Organization, countries with utility model or petty patents include Australia, Argentina, Armenia, Austria, Belarus, Belgium, Brazil, Bulgaria, China, Colombia, Costa Rica, Czech Republic, Denmark, Estonia, Ethiopia, Finland, France, Georgia, Germany, Greece, Guatemala, Hungary, Ireland, Italy, Japan, Kazakhstan, Kenya, Kyrgyzstan, Malaysia, Mexico, Netherlands, OAPI, Peru, Philippines, Poland, Portugal, Republic of Korea, Republic of Moldova, Russian Federation, Slovakia, Spain, Tajikistan, Trinidad & Tobago, Turkey, Ukraine, Uruguay and Uzbekistan. See http://www.wipo.int/sme/en/ip_business/utility_models/utility_models.htm for more information.

\textsuperscript{17}I do not argue that inventions with low sunk-costs are necessarily obvious (although the courts could usefully consider lack of sunk costs as a suggestion of obviousness just as they have used the presence of sunk costs as a suggestion of non-obviousness).
specification.¹⁸

In industries such as biotechnology that are rapidly progressing because of scientific advances there is relatively little prior art. It is therefore possible to make very broad patent claims in these industries for what are in-essence routine innovations. In a rapidly progressing industry absence of prior art may signify nothing more than that the innovator was slightly quicker off the starting block. In more slowly progressing industries absence of prior art suggests that an innovation would not have been made absent the innovators. Taking into account innovation to imitation costs can be especially important, therefore, in assessing how broad a claim should be allowed on patents in rapidly progressing industries.

A straightforward application of the economic theory would suggest that an innovator should not be allowed to claim embodiments when the imitation costs for that embodiment are nearly as high as the claimant’s innovation costs. At the time the patent is granted, however, it may not be obvious what the imitation costs are - precisely because as patent breadth increases, imitation costs become more uncertain. The Agracetus soybean patent (European) covers any genetically altered soybean created through any method but it’s quite likely that the costs of creating a genetically altered soybean through a method not in the Agracetus specification are nearly as high as the costs of using the method actually specified. Thus broad claims are suspect but the true imitation costs on the broader claims may not be known until after the patent is granted.

Since it is difficult to estimate imitation costs for broad claims at the time that a patent is granted, it may be desirable to let patent examiners allow liberal claims but with the understanding that the courts will rule on infringement suits more conservatively as cost information arrives. Courts can do this under the doctrine of “undue experimentation.” Under this doctrine “an alleged infringer can argue noninfringement by showing that extensive experimentation beyond what was disclosed in the patentee’s specification was required to make the allegedly infringing embodiment” (Merges and Nelson 1990, 849). Although monetary cost is not an explicit factor in undue experimentation doctrine, the quantity of experimentation, the required skill of the experimenters, and the predictability of the experimentation are all implicit cost factors that the courts use to decide whether undue experimentation makes for infringement.¹⁹

¹⁸It is assumed that one skilled in the relevant art can make and use embodiments that are claimed but not disclosed.

¹⁹Courts may also reject an infringement claim based on the “reverse doctrine of equivalents.” The reverse doctrine allows a literal infringement on the claims if the infringing
Prior to Edison’s invention of the lightbulb, for example, Sawyer and Mann had been granted a patent that claimed any carbonized fiber or textile material as an incandescent conductor (despite the fact that their patent only specified carbonized paper as one such embodiment). Edison argued that the Sawyer and Mann patent was too broad since most carbonized fiber or textiles would not be good conductors and the patent did not explain how to discover which of the thousands of such possibilities would be suitable. Edison pointed to his own experiments with thousands of different types of material and successfully argued that Sawyer and Mann’s patent did not lower his costs of discovering a suitable filament (Merges and Nelson 1990). In essence, Edison argued that his “imitation” costs were at least as high as the Sawyer and Mann “innovation” costs and thus their patent should not be interpreted so broadly as to cover his innovation.

Edison won his case, but a good argument can be made that excessively broad patents slowed progress in a number of other key areas such as automobile, airplane and radio technology. Unfortunately, the undue experimentation doctrine is not common and Merges and Nelson (1990, n.46) suggest that the courts have become less willing to accept undue experimentation arguments in recent decades.

Note that both the reverse doctrine and the undue experimentation doctrine can be understood as applications of the more general innovation to imitation cost principle - a claim should be patentable/enforced only if the innovation costs are high relative to imitation costs for that claim. Patent reform based upon the economic theory of patents encourages the use of the undue experimentation doctrine especially in so far as innovation to imitation costs are recognized as a key component.

Summarizing, innovation to imitation costs can be taken into account in at least four ways, 1) in deciding what product classes should receive patent protection, 2) in deciding whether a particular patent should be granted, 3) in assigning patent duration or 4) in assigning patent breadth. Although innovation is so different in principle from a claimed product or process that it performs the same or a similar function in a substantially different way. The reverse doctrine is rarely used.

The Selden patent on automobiles, for example, nearly prevented Henry Ford from producing mass-manufactured, low-cost automobiles and, as it was, involved him in nearly a decade’s worth of expensive and uncertain litigation (Greenleaf 1961). Similarly, the Wright brother’s patents slowed the industry down so much that on the eve of World War I the government, for national security reasons, stepped in to prevent patent disputes (Bittlingmayer 1988). The Marconi patent on the two-element diode tube prevented the far superior triode from being used for a long time (Merges and Nelson, 1990). See also Plant (1934), Cole (2001).
changing any of the first three would require new legislation, sunk costs can be given a greater role in assigning patent breadth via the discretion of patent examiners and minor changes in court interpretations of the undue experimentation and reverse equivalents doctrines.

5 Objections and Implementation

Perhaps the strongest objection against a patent system that takes into account sunk costs is that measuring sunk costs is difficult (Scotchmer 1988). If the sunk costs are born by the patent applicant then this problem does not seem overly difficult. In support of his own proposal to make “patents slightly easier to obtain for the results of high-cost research projects” Merges (1992) cogently argues that:

Proving relatively high research cost will not be difficult or burdensome. Patent applicants and patentees collect this information anyway for a variety of reasons, including: (1) tax benefits (e.g. the R&D tax credit), (2) internal cost accounting, (3) use in project evaluation, (4) use in licensing negotiations and the like. Patentees appear to have no trouble showing research expenditures at the damages stage of a patent infringement suit, and, as noted above, such information has been introduced in some cases to show the non-obviousness of the invention involved. Simply adding one more reason to collect data on the cost of a research project does not appear to pose a major problem.

In fact, for some currently patentable products such as computer software and business methods it may be easier to estimate sunk costs than to investigate the history of prior art or make judgements on novelty. In these sorts of cases, estimating sunk costs could reduce net patent investigation costs because a finding that sunk costs were low would preclude the necessity of investigating prior art or making judgements on novelty.

The relevant sunk costs, however, may not always be born by the patent applicant. If Thomas Edison experiments with 3000 elements for the light

\[21\] Scotchmer (1988) only devotes three paragraphs to the issue of using cost measures in the assignment of patents - which is surprising since she writes “it seems obvious that the patent system would be improved if patent values were individually chosen to cover R&D costs.” Scotchmer rejects a system based solely on costs but does not investigate the more relevant question: could the current system be improved by marginal changes that use costs as a supplementary factor?
bulb filament he can make a good case to the patent board that innovation to imitation costs are high and the light bulb is deserving of a patent in order that such costs be recouped. But suppose that there are 3000 inventors each of whom experiments with one filament. Although the innovation to imitation costs are just as high for society, the lucky inventor who hits upon carbonized cotton thread will have few sunk costs to bring before the board. If inventors experiment only because of the possibility of recouping their costs in expectation then the lucky inventor should be granted a patent despite having expended few sunk costs of his own. The patent should be granted, despite the fact that the inventor expended few sunk costs, because the lottery winnings from the patent must be large enough to motivate the 2999 inventors who experimented and received nothing.

Estimating innovation to imitation costs when some of the innovation costs are not borne by the patent applicant is certainly more difficult than estimating applicant innovation costs. But it should usually be possible to weight-up to take into account the number of investing firms. If there are \(N\) investing firms, for example, the social sunk costs could be taken to be roughly \(N\) times the applicant’s sunk costs.\(^{22}\) In most industries the number of innovating firms is few and easily determined.

In order to improve the current system it is not necessary to estimate sunk costs precisely so long as they can be estimated well enough to reasonably assign patent duration to a limited number of categories. If sunk costs justify a patent of 20 years, for example, it’s unimportant whether they are two, three or four times greater than necessary. The thrust of this proposal is to make marginal changes in the current patent system rather than to replace that system with something entirely new.

Estimating sunk costs may also be difficult when there are substantial overhead or background costs that are difficult to assign to individual research projects. Aside from the point made above that we need only enough cost information to assign a patent into a limited number of categories, there is also the more general issue that the patent system is far from the only way that research and development is rewarded. Education is the most important indirect input into R&D, for example, and it is already heavily subsidized. There are also a variety of direct subsidies for R&D. The tax code, for example, contains substantial provisions for tax credits and tax breaks that need not be assigned to specific projects nor even lead to patented or marketed products. More specific subsidies are also possible.

\(^{22}\) Note that in some models of patents races there is overinvestment in R&D thus the appropriate weight can be less than \(N\). See Tirole (1988) for an overview.
The Orphan Drug Act of 1983/1984, for example, increased $\alpha$ by increasing tax breaks on R&D for sponsors of rare drugs (see further below). In addition, governments, foundations, and charities routinely spend billions of dollars financing research directly, even when that research may also lead to patented products. Given the many policy levers that we have available we should aim to create a patent system that is efficient at the margin and use the other levers to insure infra-marginal efficiency.

Furthermore, the influence of the patent system on innovation should not be exaggerated. The vast majority of innovations in most industries would occur without the existence of patents (see section 6 below for a review of the empirical literature). It takes time for new ideas to diffuse and being the first to market, learning by doing, capturing market share when consumers face switching costs, secrecy, and other factors are in practice more important sources of competitive advantage than patents for most firms most of the time. The empirical studies indicate that with the exception of a few industries with very high innovation to imitation costs (industries for which the proposal recommends strong patent rights) we should not be overly concerned that a weakening of patents will result in underinvestment.

Under certainty an optimal patent system would assign $\alpha$ (duration, breadth, scope etc.) such that the discounted stream of per-period monopoly profits just covered R&D expenses. Scotchmer (1988) is correct that if we naively tried to duplicate this system in our uncertain world there would be little incentive to invest because some research never leads to a marketable product. It follows that ex-post profits must be positive if ex-ante profits are not to be negative. But this is not an argument against adjusting our current patent system to better take into account sunk costs. Any patent length short of infinity deters some research and development. We accept this possibility because the deadweight losses and other costs of extending the patent system exceed the benefits of the additional research that would be generated. But, even after we have recognized this tradeoff, it would be remarkable if 20 years were the optimal duration regardless of the size of sunk costs. In other words, we should not try to adjust patent duration to eliminate ex-post profits but we should try to adjust patent length so that the tradeoff between generating valuable research and enduring deadweight losses is optimized on as fine a level as is practical.

23The higher are sunk costs the less likely it will be that the “natural” barriers to entry will last long enough for inventors to recoup their sunk costs. Thus, once again, higher sunk costs justify stronger intellectual property rights.
The best evidence that it is feasible to take into account sunk costs in the assignment of patents, in a limited yet useful way, is that we have done just this in one prominent case, the Semiconductor Chip Protection Act (SCPA) of 1984.24

Prior to 1984, computer chip designs were not generally patentable or copyrightable. In 1984, however, Congress created a new form of intellectual property, mask works. (The mask in essence is a set of stencils that is used to lay down the topography of a micro-chip). Owners of mask works are given rights that last for a period of 10 years but that are in certain respects weaker than patent rights. Reverse engineering a chip in order to produce another chip, for example, is legal in certain circumstances. In particular:

If the resulting semiconductor chip product is not substantially identical to the original, and its design involved significant toil and investment so that it is not a mere plagiarism, it does not infringe the original chip, even if the layout of the second chip is, in substantial part, similar25

Thus the SCPA establishes a process whereby a new chip design does not infringe an old design, even it is similar, so long as substantial toil and investment was needed to create the new design. In other words, the standard for protection of the new intellectual property is that imitation costs be substantial relative to innovation costs (this standard can also be thought of as a codified version of the undue experimentation doctrine discussed earlier). The substantial toil and investment standard is consistent with the economic theory of patents because it prevents imitators from free-riding on the initial design but nevertheless does not prevent competition from non-free riders. Interestingly, Congress foresaw the objection that the substantial toil and investment standard would be too difficult to implement but they rejected that objection noting that the test is not “unduly difficult” because the additional work required to establish the privilege will ordinarily leave a paper

24 Another minor but telling piece of evidence is that costs were also taken into account in the operation of the aircraft patent pool. New patents added to this pool were to be royalty free unless the patent “secures the performance of a function not before known to the art...or the amount expended in developing the same is such as to justify such compensation.” Quoted in Bittlingmayer (1988). Note that a patent pool is a private, contractual substitute for the government patent system. Thus, it is telling that private actors should use costs to decide on royalty rates.

Note that in addition to taking into account sunk costs the SCPA also created intellectual property protection of duration less than 20 years, thus demonstrating that at least on a product class basis the intellectual property system can handle patents of varying duration. Another example of varying patent protection by product class occurred in the Orphan Drug Act of 1983/1984. The Orphan Drug Act increased intellectual property protections for rare diseases, defined as diseases affecting fewer than 200,000 people. In addition to extending generous tax breaks on research and development, the ODA gave sponsors of rare diseases seven years of market exclusivity, meaning that for seven years, no other company could obtain permission to market a similar drug (Tabarrok 2002). The seven years of exclusivity is available in addition to any patent protection.

6 There is Little to Lose from Patent Reform

If it were clearly true that on net the patent system increased economic growth and technological advancement then on precautionary grounds alone there would be a good case against reform. But the consensus from many studies of innovation is that most innovations would occur without patents. Surveys of innovations by Mansfield et. al. (1981), Mansfield (1986), Levin et al. (1987), and Cohen et al. (2000), for example, all suggest that in most U.S. industries patent protection has not been an important encouragement of R&D and has not been essential to innovation. Mansfield (1986), for

---

26 Evidently Congress wanted to avoid some of the problems of patent law that have been created by the patenting of procedures and ideas and noted explicitly that “In no case does protection under this chapter for a mask work extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery regardless of the form in which it is described, explained, illustrated, or embodied in such work.” 17 U.S. Code 902(c). See Kasch (1992) for more on the SCPA.

27 Creating a drug and getting it approved by the FDA to treat a rare disease can be just as costly as creating a new anti-cancer drug. Yet the potential profits from a drug treating say Gaucher’s disease are much smaller than those for an anti-cancer drug. Recall that in an optimal system high value property is weakly protected and low value property strongly protected. The ODA recognized this by creating stronger intellectual property protections for pharmaceuticals that combat diseases that affect a relatively few patients.

28 In an exhaustive survey of over 150 years of data from some 60 countries, Lerner (2002) finds that increased patent protections result in only negligible, and perhaps negative, increases in patenting by inventors in the country increasing patent protections - a seemingly minimum standard for an increase in innovation. Indeed Cohen et al. (2000) find that much patenting is simply generated by the patent system itself as firms patent in order to protect themselves from the patenting of other firms (see also Hall and Ziedo-
example, examines detailed data from 100 firms in 12 industries and concludes that in most of these industries patent protection was essential for the development of less than 10 percent of the innovations. In only two industries, chemicals and pharmaceuticals was patent protection necessary for a third or more of the innovations. If patents are necessary for only a minority of innovations then on the vast majority of innovations the monopoly distortions created by patents are unredeemed.\(^29\)

Chemicals and pharmaceuticals, the two clear outliers, are perhaps the best examples of industries with high innovation to imitation costs so the fact that patents have been important to creating innovations in these industries is consistent with the economic theory. It is reassuring that the reform proposals given here would independently approve of patents in the area where patents appear to work best.

I have motivated patent reform by reference to economic theory but the empirical literature on patent effectiveness provides an equivalent motivation. Patents should be strong when they are effective in generating innovation and weak otherwise. The empirical evidence indicates that this means patents should be strong when innovation to imitation costs are high and weak otherwise.

Given the weak arguments that can be made on behalf of patents the arguments for reform cannot be rejected using the precautionary principle.

7 Other Theories of Patents

The patent system does not follow the logic of the economic theory of patents (certainly not to the extent that it could follow that logic given that any of the above reforms are feasible). If the patent system doesn’t fit with the theory either the patent system needs fixing, the route I have emphasized, or the theory is wrong (or at the very least incomplete). Could the patent system be better explained, and perhaps shown to be optimal, according to another theory?

\(^n\) nis (2001) who find that wasteful defensive patenting is the norm in the semiconductor industry).

\(^29\) Recent research suggests that in addition to creating monopoly distortions patents can actually reduce innovation. Bessen and Maskin (1999), for example, show that patents can reduce social welfare when innovation is sequential and complementary. Briefly, in any industry that builds innovations on previous innovations, patents can preclude the imitation that is necessary for progress. See also Hunt (2001) and Hunt (1999) for models and evidence suggesting that patents can reduce innovation. See Cole (2001) for a survey of the arguments against patents.
After the economic theory, the disclosure theory of patents is probably the most common although it has not been well received by economists (Machlup 1958). Patents are said to be the price that must be paid to induce the innovator to disclose the workings of the patented product or process, thus allowing for more rapid diffusion of the underlying knowledge. If anything, however, the disclosure theory seems more at odds with the patent system than the economic theory. The disclosure theory, for example, cannot explain why patents are allowed on products where disclosure is made self-evident by the product itself. The workings of many mechanical innovations and tools, for example, are evident on inspection and thus on the disclosure theory such tools should not be patentable. Even more oddly, many products, particularly pharmaceuticals and other chemicals, can easily be reverse-engineered. Thus the disclosure theory would suggest that pharmaceuticals should not be patentable!

More generally, the disclosure theory suggests that patent duration (or other factor influencing α) should be a function of the likelihood of non-disclosure absent the patent. Intellectual property should be more greatly protected, for example, when trade secrecy is a relevant alternative to patenting. The patent system, however, makes no attempt to evaluate or take account of the likelihood of non-disclosure absent the patent.

It may be true that the patent system occasionally encourages disclosure and that this is to be counted as a benefit of the system. It does not seem plausible, however that the patent system is designed to optimize the benefits of disclosure nor does it seem desirable that the patent system should be modified along the line suggested by the disclosure theory. The disconnect between the economic theory of patents and the patent system cannot, therefore, be explained on the grounds that the patent system is optimal in the disclosure dimension.

Space precludes a detailed examination of other utilitarian theories of patents. If it is difficult to reconcile utilitarian theories of patents with the patent system, however, perhaps this is because the patent system has been influenced by non-utilitarian theories. Non-utilitarian, “natural-rights,” justifications for patents, for example, were common at the time of the founding of modern patent systems especially in Europe (see Menell 2000 for more on non-utilitarian patent theories). The French Constitutional Assembly, for example, in the preamble to the patent law of 1791 stated that:

---

30 The patent system has also been accused, however, of delaying disclosure particularly of research that is done at universities. University research of a commercial nature would often be disclosed earlier were it not for the necessity of keeping the information secret until a patent is obtained.
...every novel idea whose realization or development can become useful to society belongs primarily to him who conceived it, and that it would be a violation of the rights of man in their very essence if an industrial invention were not regarded as the property of its creator. (Qtd. in Machlup 1958, 22).

The (possible) influence of non-utilitarian theories on the patent system can explain one aspect of patent systems that is especially peculiar from the perspective of utilitarian theories - the fixed duration of patents. The natural rights of a poor farmer to a small plot of land are no different than those held by the owner of thousands of acres - in the same way it makes sense, given the natural rights theory, that every inventor should hold the same rights to her invention regardless of invention quality, cost, potential for secrecy or other factor. Of course, the natural rights theory would suggest, contrary to practice, that patents should be held in perpetuity - this position was indeed strenuously argued in Europe at the time that modern patent systems were founded.

Since utilitarian and non-utilitarian theories of patents begin with different premises it would not be surprising that they reach different conclusions. Thus, if non-utilitarian theories influenced the founding of patent systems the tension between utilitarian theories and some aspects of patent systems is to be expected. The importance of this is that we need not accept the ‘what is, must be efficient’ viewpoint. Utilitarian reforms can, by definition, raise the utility of non-utilitarian systems.

8 Conclusion

Patents are justified in the standard economic theory when innovators must incur substantial sunk costs that need not be incurred by imitators. The theory suggests, therefore, that the relative cost of innovation to imitation should be a key consideration in deciding what particular products or what sorts of products deserve patent protection. Yet patent law pays little attention to costs. Patent law is divorced from the economic theory of patents. As a result, patent law extends protection to many classes of innovations - including software, medical procedures, and business processes as well as to many vague “ideas” - where innovation costs are low relative to imitation costs. Unnecessary monopoly distortions and perhaps reduced innovation are the result. A patent system based on the economic theory of patents would take sunk costs into consideration - extending fewer protections when
sunk costs are low. Sunk costs could be taken into consideration in the allowing of patentable product classes, in ruling on individual patents, in setting the duration of patents or in setting patent breadth. Of these, changing the breadth of patents could be most easily accommodated within the current system. Recent changes in patent-related law suggest that changes along these lines are feasible even if they have not yet been implemented.

Acknowledgments: I thank Bruce Benson, Bryan Caplan, Tyler Cowen, Eric Helland, Gordon Tullock and two anonymous referees for comments. I thank Aaron Edlin for especially useful comments.

References


