
Remarks prepared as a member of the Economic Advisory Group,
European Patent Office, Munich
December 10, 2004

Suzanne Scotchmer
2607 Hearst Avenue
University of California
Berkeley, CA 94720-7320
scotch@socrates.berkeley.edu

¹I thank Rochelle Dreyfuss and Dominique Guellec for helpful discussion.
1. Preface

The patent system has recently been under attack by policy commentators who question whether it serves the incentive purpose for which it is intended. Such commentary is sometimes provoked by patents such as the famous one-click patent of Amazon, Inc., which seem unnecessary for any incentive purpose, and sometimes by whole subject matters such as computer software, for which it is uncertain whether patents are necessary, or gene sequences, which may not satisfy the legal requirements for a patent. In these comments, I will try to disentangle the several types of criticisms currently directed at the patent system, and comment on their economic merit. I distinguish between patent quality and patent design. The arguments and evidence summarized here lead to the conclusion that procedures for administering the patent system have an important impact on design as well as quality, and much of the current controversy concerns both.

In Europe, the fusion of quality with design is even more complicated because national patent offices and the EPO can both make policy. Given that administrative procedures affect the design of intellectual property rights (especially the examination function), the division of responsibilities between the EPO and national patent offices should have substantive consequences. As a framework for considering the consequences, I summarize arguments for why, within a system of national treatment, nations may disagree on appropriate intellectual property protections, and why nations may try to modify their policies to suit their national interests within the constraints imposed by TRIPS.

2. Patent Quality

An excellent review of what we know about patent quality in the U.S. system can be found in the very recent (2004) report of the National Academies of Sciences, together with the background research documents (NAS 2003). This paper gives a
slightly different focus to their discussion by distinguishing between patent quality and patent design, and by calling attention to the important role of patent breadth as a policy lever, rather than focussing on inventive step, as the NAS report does.

Patent quality generally refers to whether the claims granted in a patent meet the legal requirements, in particular, appropriate subject matter, utility, disclosure, enablement, novelty and nonobviousness (sufficient inventive step). The difficulty is that these standards must be interpreted anew for every new technology. On one hand, the applicant will generally be insisting on the most generous interpretation, and on the other hand, the courts are the ultimate arbiters of what is permissible. The patent office must anticipate the outcome of litigation or of a challenge such as re-examination or opposition, and where there is no clear law, make policy, as in the 2001 *Utility Examination Guidelines* they issued for genetic sequences. Due to the administrative rules (Rai 2004), the patent office therefore assumes some responsibility for design as well as implementing the law. This makes the boundary between quality and design fuzzy.

The difficulties for patent examiners are compounded by their workload. The number of examiners per 1000 applicants to the USPTO has fallen by about 20% in the past five years, resulting in longer delays to issuance (24 months, up from 18 months), but without reducing the time devoted to each examination (a little less than 20 hours per application (NAS 2004, p. 51)). The charge has often been made that it is easier for an examiner to resolve ambiguities in favor of the applicant, since that allows the examiner get on to the next application. Compounding these pressures is the presumption in the American system that the applicant has a right to the patent; the burden of proof for rejecting the application is on the patent examiner, and there is a presumption of validity in a litigation.

The PTO itself is concerned with errors. It annually re-examines two to three percent of issued patents, and finds an error rate that has fluctuated between 3.6 and 7 percent since 1980, recently being about 4% (NAS 2004). Another indicator of
patent quality is the re-examination rate. About 0.3 percent of U.S. patents are re-examined, with about 10 percent revoked, and two-thirds modified (Graham et al 2003). However, it is not clear that these changes are due to “error”, since close to half of re-examinations are proposed by the patent holder himself in light of newly discovered prior art. The NAS also reports that in recent years the rate at which the EPO and JPO approve patents has been lower than the rate at which the USPTO approves patents (NAS 2004, figure 3-2), possibly suggesting that the USPTO grants patents more liberally than their colleagues elsewhere.

The suspicion of low-quality patents has surfaced in the context of expanding rights in the U.S., in particular to computer software and business methods in the 1990s. For examples and discussion, see Barton 2000 and Lemley 2001, writing in large part about business methods implemented with computer software. Since patents on computer software emerged rather precipitously, and business method patents even moreso, most of the prior art was not in the main database consulted by patent examiners, which is prior patents. Instead the prior art was largely to be found in industry practice, in existing computer programs, and to some extent in academic publications. All of these are harder to search than prior patents. The NAS echoes the criticism of many others in suggesting that these search problems have resulted in software patents that should have been barred by prior art. In the case of business method patents, some of the ambiguity as to design was resolved by Congress rather than the PTO, when Congress enacted the First Inventor Defense Act of 1999 creating exemptions for prior users.

Perhaps because PTO proceedings tend to be invisible, requests by rivals for re-examination are rare, especially as compared to European oppositions. This is even though oppositions must be filed within nine months of issuance, as opposed to re-examinations, for which there is no time limit. In addition to being more common, European oppositions more often result in modification or revocation of the patent (Graham et al 2003).
Few observers doubt that patent offices sometimes issue patents in error, either with respect to their own guidelines, or in the sense that a court would eventually disagree. What is less obvious is the harm that such patents do, given that errors can be remedied by courts, re-examinations, and by voluntary licensing among parties. The harms listed by the NAS (2004, p. 95) are

- In contrast to incentives to genuine innovation, patents on trivial innovations may confer market power or allow firms to use legal resources aggressively as a competitive weapon without consumer benefit.

- Poor patents could encourage more charges of infringement and litigation, raising transaction costs.

- The proliferation of low-quality patents in a technology complicates and raises the cost of licensing or avoiding infringement.

- The uncertainty about the validity of previously issued patents may deter investment in innovation and/or distort its direction.

It is clear from this list that the boundary between quality issues and design issues is fuzzy. The first problem can arise either because the patent office issues patents on unpatentable inventions, or because the standards for patentability are too low as a matter of law. The third problem, of patent proliferation and financial burdens due to licensing, can arise even for legitimately issued patents, depending on the standard for patentability. The second harm is linked to uncertainty about the legitimacy of issued patents, but uncertainty is not the same as low quality. If bad patents are understood as such by the patent holder and rivals, then the potential litigants have an incentive to avoid litigation by licensing on terms that reflect the likely outcome of litigation. If there is real uncertainty about the law, that problem may or may not be due to the patent office. Although the patent office has the right and responsibility to implement consistent practices, the ultimate arbiter of design are courts. Like the second problem, the fourth problem also reflects lack of clarity.
in the law. If it were clear how previously issued patents would be resolved in court, then earlier patents would not impinge improperly on future innovations. If the earlier patents were improperly issued, the later innovator would predict the outcome of a judicial proceeding and not be deterred.

Admittedly, these arguments assume that disputes with predictable outcomes are cheap and easy to resolve. That may be false, but the argument highlights an important difference between patent-office error and lack of clarity in the patent law. The natural meaning of “low-quality patent” is that everyone agrees it was an error. If so, potential litigants will achieve by licensing or other means the correct outcome (the anticipated outcome of litigation) without paying litigation costs. On the other hand, true uncertainty about the patent law (or, equivalently, uncertainty about the outcome of litigation) often cannot be avoided by diligence at the patent office.

This is an area we know little about empirically. We do know that litigating patents in the U.S. is expensive, between one million and four million dollars for disputes with damages between one million and twenty-five million dollars (NAS 2004, p. 70), but the litigation costs apply only if the dispute gets litigated. If a dispute settles on terms that reflect the likely outcome of the court proceeding, the harm inflicted by a bad patent is limited. Nevertheless, due to the inherent vagaries of legal process and the fact that litigation costs create holdup power, a bad patent can inflict harm to users even if everyone agrees that the patent was issued in error. Bad patents should certainly be avoided if it is costless to avoid them.

Since the harm due to bad patents is unclear, it is also not clear whether high examination costs should be incurred to avoid them. There have been arguments on both sides. Merges (1999) argues for better examination on the ground that it is cheaper to resolve potential conflict or to remedy errors at the examination stage than at the litigation stage. But Lemley (2001) reminds us that more patents are examined than litigated. Lanjouw and Schankerman (2004) estimate that between 1% and 2% of patents are litigated, mainly the high-value patents. If a court
proceeding is viewed as an elaborate “examination,” then the costs of examination are only incurred for a small percentage of patents. It is worth noting, however, that both the report on patent reform of the U.S. Federal Trade Commission (FTC 2003) and the 2004 NAS report endorse an increase in funds for patent examination.

3. Patent Design

The previous section pointed out that patent quality is an imprecise concept that shades into patent design, and that patent offices have influence on design as well as quality, especially in the breadth of claims that they grant. In this section I turn to design issues, focussing on breadth as a policy lever available to the patent office, rather than inventive step, which is the focus of the NAS report.

A deep conundrum of patent law is that the incentive purpose – to motivate R&D investments when the expected social value outweighs the cost – is not articulated in the statute, at least in the U.S. patent law, 35 U.S.C. Nowhere does the statute or case law say how the requirements for a patent are supposed to relate to incentives. The absence of guiding principles as to the economic purpose leaves judges free to make up their own. In 1941, Justice William O. Douglas famously declared that for an invention to be patentable it must reveal a flash of creative genius. Even though this standard was superseded by the 1952 U.S. patent statute, it is instructive as to how legal reasoning can go astray when divorced from its economic purpose. If the invention requires only a flash of creative genius and nothing else, then what incentive role does the patent system play? If the flash of creative genius must be accompanied by an investment, why does it matter that there was a flash of creative genius? Under the 1952 statute, the invention is not patentable if it would have been “obvious” to “a person having ordinary skill in the art.” This may improve on the “flash of creative genius” standard, but can still lead to judicial rule-making that is not very tightly linked to economic objectives.
Inventors only need protection if invention is costly; the extent of protection should in some way reflect the expected costs. Further, the profit available through intellectual property protection should be given in a manner that tries to minimize the burden on users, usually called deadweight loss. The optimal design of intellectual property thus involves two related questions: What is the appropriate level of profit, and how should the profit be structured? I will comment on these two questions as they relate to the debate about patent quality. (My 2004 book gives a more complete treatment.)

The main policy levers of intellectual property, as conceived by economists, are length, breadth and the required inventive step. As a legal matter, the potential for infringement is determined by the claims that are granted, and in the U.S., by the doctrine of equivalents. Claims can be broad or narrow, depending on how many substitute technologies the claims exclude from the market. Patent breadth is an economic concept which does not precisely reflect any particular legal doctrine. When courts strike down patent claims, they are often narrowing the patent in the sense understood by economists, as well as disagreeing with the patent office’s assessment of what inventor invented.

Barton (2003) gives an amusing example that illustrates the subtleties of the two policy levers, breadth and required inventive step. He describes a series of patents that have issued on insulating sleeves for paper cups – those convenient devices that protect your fingers when carrying coffee in a paper cup. He describes how minor changes in the pattern of dimples stamped into the cardboard sleeve have led to new patents. Barton views the patents as illustrating that the nonobviousness standard (or required inventive step) is too low, and that such patents should not issue, perhaps on an economic theory that such inventions are not very costly and will be made even without patents.

My own lesson from the example is somewhat different. One has the impression that the patents are noninfringing. If the insulating sleeves offered by different
vendors are noninfringing, the patents do not harm competition – it is as if no patents had issued at all. Such patents may be an affront to common sense, but if each is sufficiently narrow, in the sense that close substitutes do not infringe, where is the harm? Narrow claims are a powerful policy lever for reining in patents on inventions that embody some small inventive step, but do not need large rewards.

Contrast with U.S. patent 4,736,866 (1988) on the Harvard oncomouse, that transgenic creature created to be cancer prone so that it is a good research animal. Instead of claiming only the process of transforming the fertilized egg, or the specific mouse line itself, the U.S. patent claims all “non-human transgenic mammals” created using the patent holders’ method. Under such broad claims, later researchers may not be saved from licensing or infringement even if their own efforts to create a transgenic walrus require large costs and extraordinary expertise, and even if the walrus is used to study obesity rather than cancer.

Is this a problem? That depends on whether such a broad patent, or large reward, is necessary. While it would be difficult to use the breadth of patents to keep rewards commensurate with cost in every instance, it is not unreasonable to consider the costs of inventions in large classes. If biomedical research tools like the Harvard oncomouse are very costly on average, then the broad claims may be appropriate.

An obstacle for many lawyers is that the economists’ notion of breadth has no direct analog in law. The claims in the patent are limited retrospectively (with respect to existing knowledge) because the claimed invention must be distinguished from prior art. The claims are limited prospectively (with respect to future discoveries) according to what is enabled and disclosed in the patent. These limitations, which economists call “breadth” (see Scotchmer 2004a, chapter 4) are distinct from the question of whether the invention should receive a patent at all. The distinction is important in the policy debate, because narrowness of the claims can limit the holdup value of the patent. For example, the patents on insulating sleeves for paper cups seem less threatening to competition when one notices that
the patents are very narrow (noninfringing).

The same observation applies to business-method patents. Critics claim that many business method patents should be rejected on grounds of prior art. However, making the case for rejection is difficult because the prior art is so pervasive that no one would bother to write it down (NAS 2004). For example, U.S. patent 5,794,207 describes a method for conducting a Dutch (descending-bid) auction, a selling technique that has been in use for centuries. Since the Dutch auction was well known, why did it merit a patent? A legal response may be that the use or implementation of the Dutch auction meets the requirements of patent law (if only in the U.S.), and no further inquiry is necessary. Another possibility is that the the patent is truly “low-quality” in the sense that the invention does not defeat prior art. An economic inquiry would ask something different, namely, whether the patent was necessary to elicit the implementation. If not, the patent serves no economic purpose, and harms users by imposing high prices. If, on the other hand, the implementation of the Dutch auction was costly and innovative, then a narrow patent on the implementation (but not all implementations) might be exactly appropriate. The patent holder will earn some profit, but much less than if he could collect royalties from all other firms that implement a Dutch auction in a different way. Perhaps more importantly, users will be protected from high prices. This is a question of breadth.

Can these economic considerations be introduced to patent law? The biggest obstacle is one-size-fits-all. As we have mentioned, the same legal doctrines apply to all subject matters covered by the statute, and given that no economic purpose is articulated, neither patent offices nor courts have a mandate to tailor their treatment to the protection that is required. The NAS (2004) counsels against any major disruption to this system. Nevertheless, legislative bodies occasionally cleave off subject matters for *sui generis* treatment, such as the U.S. Semiconductor Chip Protection Act of 1984 (10 years of protection outside the patent system) and the European Database Directive of 1996 (15 years of protection).
In addition, if authority over design issues is *de facto* the joint responsibility of patent offices, legislative bodies and courts, then it is not inconceivable that patent offices could consider costs in their interpretation of “nonobviousness” when granting patents. This would be a major step, however, possibly causing further delays and disputes. As patent law is now constituted, all the evidence in the prosecution process is on the nature of the invention, for example, distinguishing the invention from prior art, and describing what was invented. Evidence about costs would be even more problematic. Rewards should not be based on accounting cost, as that might reward inefficiency. Overhead costs must be apportioned among projects; who would decide how to do that? And how would costs be accounted when there was an *ex ante* probability of failure? The innovator must be overrewarded in case of success to account for the probability of failure. These seem like disabling problems from an implementation point of view. Nevertheless, the mere recognition that costs should matter to the size of the reward gives some guidance as to how intellectual property law should be structured. If business method inventions are not very costly on average, then that is an argument for giving narrower breadth than for subject matters that are costly on average.

Much of the economics literature on breadth does not view breadth as a means to increase or decrease the reward to innovation, but rather addresses the structural question of how the reward should be given. Holding the total profit fixed, the same reward can be given with a narrow patent of long duration or with a broad patent of short duration. This question has been studied in the context of differentiated products (Klemperer 1990), process patents or other contexts where different rightholders may sell an identical commodity (Maurer and Scotchmer (2002), building on Gallini (1992) and Gilbert and Shapiro (1990)). The main conclusions are that it is generally better to give longer patents of narrower breadth. The narrow breadth reduces consumer loss in each period (by reducing price) more than it reduces profit.

The argument is subtler in the context of so-called “quality ladders,” where
innovators successively invade the market with improved products. Breadth is important for additional reasons than in the static context. With successive market dominance, the effective life of a patent will generally be shorter than the statutory life, due to the introduction of a noninfringing product that replaces the previous one (O’Donoghue, Scotchmer and Thisse 1998). The relevant policy variable for the duration itself is the breadth of the patent. If, for example, all improvements except radical ones are infringing, then the patent holder can expect a long period of market dominance, possibly taking licenses on small improvements from other innovators. Accompanying the longer period of market dominance is a greater market advantage, since the market incumbent will typically have a larger quality advantage over the rival he replaced. Thus, increasing the breadth of the patent increases both its profitability in each time period (due to diminished competition between successive noninfringing products) and prolongs the period in which profit is earned. The policy question is, To what extent should this advantage be given to each incumbent? The answer should clearly refer to the tradeoff between encouraging quality improvements and keeping prices under control, and that tradeoff cannot be assessed without reference to expected costs of making improvements.

It is in the cumulative context where blocking patents may arise. Patents are blocking for a particular innovation if a previous patent holder and the patent holder on the new innovation can both keep the innovation from being used or sold. Blocking patents may arise when an improving innovation infringes a prior patent, while the innovator receives a separate patent on the improvements. Blocking patents make it clear that the legal requirements of novelty and nonobviousness (inventive step) govern the two distinct policy levers, breadth and the required inventive step. Whether or not the improved product infringes the prior patent is determined by the breadth of its claims. Whether the improved product receives a patent is generally independent of this, and depends on whether the improvement is nonobvious or achieves the required inventive step. (See chapter 5 of Scotchmer 2004a for more discussion of the economic consequences of blocking, and Lemley 1997 for a discussion of how copyright law and patent law differ in this regard.)
4. Patent Politics

The instrumental view of the patent system is that it exists to promote innovation. However, there are at least two reasons that longer or broader patents may not be in the public interest. First, longer or broader patents typically increase deadweight loss to consumers on innovations that would be forthcoming even with weaker rights (Nordhaus 1969), and second, when innovation is cumulative, longer or broader patents may create obstacles for future innovators (Scotchmer 1991, Merges and Nelson 1990). Much of the economics literature is aimed at sorting out how innovation can be stimulated while minimizing those losses. Because giving strong patent rights to an early innovator may reduce incentives for later innovators, strengthening patents may not be the same thing as strengthening incentives. For more on this paradox, and how it depends on licensing opportunities, see chapter 5 of my 2004 book.

The efficiency criterion that underlies most economic analysis of patent design is focused on consumer benefits. The only rationale for patents is that they encourage development of new products, and thus generate consumers’ surplus. The net profit that accrues to inventors is also a social benefit, since it is a transfer from consumers, but profit is recognized as a necessary evil, since the flip side of profit is deadweight loss. There is no efficiency rationale for protecting inventors beyond what is required to encourage innovation.

This reasoning gets subverted in the international arena. To a trade policy negotiator, profit earned abroad is unambiguously a good thing, and the consumers’ surplus conferred on foreign consumers does not count at all. There is a domestic interest in capturing profit abroad, and symmetrically, there is a domestic interest in trying to ensure that domestic consumers get access to foreign inventions on competitive terms. That is, domestic policies toward intellectual property will be driven by profit flows, rather than (or in addition to) the effect on innovation.
In a treaty context such as the Paris Convention, where nations are obligated to provide national treatment to foreign inventors, but protections are not harmonized, profit flows can cause a “race to the bottom.” By weakening domestic protection, a country can stem the outflow of profit to foreign inventors without affecting the inflow of profit to its own inventors from foreign consumers. But all domestic policy makers realize this, and if they all weaken protections accordingly, the incentives to innovate may be undermined. For example, in a two-country system, if one country ceases to protect a given subject matter, the other country may reciprocate. Otherwise, the country providing unilateral protection would end up providing the rewards to all innovators, both foreign and domestic, while the new technologies or creative works could be used in the foreign market at competitive prices. Adding insult to injury, the foreign innovators would not even collect proprietary prices in their own countries. (For a more systematic treatment of when and whether a race to the bottom is a threat, see Scotchmer 2004b or chapter 11 of Scotchmer 2004a.)

TRIPS went beyond national treatment, and required members to harmonize their protections. One interpretation of TRIPS is that it attempts to rescue the global innovation system from the counterproductive incentive to free ride. For the most part, TRIPS obligates countries to provide certain protections, but does not prescribe stronger protections. Thus, the treaty only works in one direction, the direction of more protection.

But can harmonization go too far? This is a difficult question, since it is not clear what the appropriate benchmark for global optimal protection is, and since countries with different characteristics will disagree. In general, the countries that should prefer stronger harmonized protection are those that are more innovative and those that are smaller. A domestic cost-benefit analysis compares deadweight loss in the local domestic market against profits that its innovators can earn abroad as well as at home. The comparison is more favorable to strong protection if the domestic market is small and the country is very innovative. In that case, the additional deadweight loss of stronger global protections would accrue to only a few domestic
users, but that loss would be balanced against a large inflow of profit from abroad (Scotchmer 2004b).

Harmonization gives nations a means to avoid the “race to the bottom.” Nevertheless, once the harmonized protections are in place, the logic above suggests that countries may still want to reduce their own protections. The same dynamics may be at work in the smaller context of the European Patent Convention as in the global context of TRIPS. Even if the only remaining flexibility is at the administrative level, the forces outlined above may have predictive value as to how that flexibility will be exercised. I return to this question after commenting further on the locus of authority regarding both patent quality and patent design.

5. Who Makes Patent Policy?

The 2004 NAS report does not make a clear distinction between harms that can arise from low-quality patents and harms that can arise from badly designed patent law. As I have tried to argue above, this is because quality and design are fused through the shared responsibility of Congress, the PTO and courts for both. In particular, it is not true that Congress and the courts are solely responsible for design, while the PTO is solely responsible for issuing high-quality patents.

The NAS makes recommendations of two types. Some are about substantive law (“reinvigorate the nonobviousness standard”) and some are about procedure (“institute an open review procedure”, “reduce redundancies and inconsistencies among national patent systems”). Because of the fused responsibilities, it is not always clear to whom their recommendations are addressed. This question of responsibilities may well turn out to be the next frontier of patent debate, which Arti Rai (2004) has already broached. In particular, the U.S. Administrative Procedures Act requires that courts give deference to administrative agencies in adjudicating substantive disputes. In 1999, the Supreme Court clarified that this
principle extends to the PTO in its relationship to the Court of Appeals for the Federal Circuit, which adjudicates most patent disputes at the appellate level, including those between an applicant and the PTO itself.

The force of all this is that policy making is partly an administrative function. The EPO adds the politics of TRIPS to the administrative complexity of any patent system, and this should be taken into account in designing the administrative relationships between national offices and the EPO. In the same way that TRIPS binds member states against a race to the bottom, the EPC might want to bind the administrative practices of member states, for the same purpose.

Above I stressed breadth as an important element of patent design which is at the heart of many patent controversies, and mostly under the control of examiners. Breadth is most immediately governed by the claims that are granted. Examination therefore has a large impact on patent design, even if the prior question of patentability (inventive step) has been resolved.

In the European Patent Convention, national patent offices will have more control of patent design if they, rather than the EPO, perform the examination function. As a corollary to the logic in the previous section, we might expect those nations that favor strong protections to be most in favor of EPO examination, since examination by the EPO binds other nations from chiseling away the strength of protection. The arguments for harmonized protections under TRIPS may also be arguments for examination in the EPO rather than in national patent offices. Of course this also works in reverse: Some scholars argue that TRIPS goes too far in its harmonized protections, and that such “chiseling” is therefore in the public interest.
REFERENCES


