ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 663 December 2008

Property Rights and Invention

Katharine Rockett

Property Rights and Invention

December, 2008

Katharine Rockett

University of Essex, University of Tromso, CEPR

Preliminary. Comments Welcome

Draft chapter 7 for Handbook of Economics and Technical Change, Bronwyn Hall and Nathan Rosenberg, Editors. This chapter was partially sponsored by "The Knowledge-Based Society", Research Council of Norway project 172603/V10.

I. Introduction

This chapter presents a selective survey of the literature on the theory of intellectual property and innovation incentives. We will outline the broad effects that have been identified with an eye to presenting a good range of modelling styles and issues. With regret, length considerations mean that we only cover patents.

We start by sketching the patent right and the associated institutional background that we will refer back to later when discussing the models. Next, we pass to some broad modelling issues, including the goals of the patent system and how they have been translated into objectives for policymakers. We outline basic theories of intellectual property design and the welfare tradeoffs they suggest. As a benchmark, we briefly consider how far we can get towards our social goals without intellectual property protection or, alternatively, how the incentive to innovate can be conceptualised when actors opt for secrecy as a means of protecting intellectual property. Next, we will look at the issues in patent design that have been investigated over the last forty years or so. We start with the simplest case of a single, one-time innovation and then look at how our conclusions on design change when we consider innovations that either build on each other or fit together as complements. We start from models that consider a system of intellectual property protection that is quite similar to existing institutions, then move on to mechanisms that start from something closer to a "blank slate" to optimally procure innovation. We consider issues of enforcement and the interplay between competition policy and intellectual property policy briefly at the end of the survey.

II. A Brief Sketch of the Patent Right

In order to set the stage, we elaborate here some of the salient features of the patent right that underlie the models that follow[1]. The US and Europe will be our focus. Indeed, the differences between the US and Europe illustrate the range of policy tools that can be brought to bear on patent design. Further, the differences add up to a somewhat different patent right on the two sides of the Atlantic, with a tighter, more expensive, and more industrially-oriented version in Europe.

What is a patent?

A patent[2] refers to a temporary property right on an invention. The patent provides a right – but not a guarantee -- to exclude others from making, using or selling the patented property. Indeed, the patent holder generally has no obligation or necessarily even the right to practice the innovation. For example, if inventor A is granted a patent, where the exercise of that patent would infringe the patent rights of inventor B, inventor A has no automatic authorisation to exercise her patent. Her right is

[1]A more complete description of the patent right, including its history, can be found in Scotchmer (2004), Guellec and van Pottelsberghe de la Potterie (2007), and Jaffe and Lerner (2006).

[2]See Guellec and van Pottelsberghe (2007) and Jaffe and Lerner (2006) for more detailed discussion of the philosophical differences between European and US views of patents. For an argument for a constitutional underpinning of patent rights, see Nard and Morriss (2006).

dependent on the patent rights of B[3]. Furthermore, the exercise of patent rights must fall within other laws, such as antitrust laws. In the same way as an individual may own a gun but not be allowed to shoot people with it, an inventor may be granted a patent but not be allowed to use it entirely as she pleases: she must do so within the scope of the general body of law.

In exchange for these exclusionary rights, the patent holder must publicly disclose the invention as part of a publicly available patent document. In the US, this public disclosure must be such that a person with sufficient background knowledge ("skilled in the art") could make or use the innovation in its "best mode" at the time of filing. Of course, for patents filed relatively early in the development cycle the "best mode" may be quite rudimentary, lacking many of the improvements that make the invention economical to exploit. While there are some differences in how it is interpreted between Europe and the US[4], the disclosure should be viewed as broadly helpful to third parties wishing to understand the nature of the innovation. While the embodiment of the innovation is protected by the patent, the underlying idea is not. Furthermore, the idea should be – and generally is -- relatively transparent in the disclosure.

The features of the innovation must be described in a set of claims, which define the metes and bounds of the patent. Features not claimed are not covered by patent protection. While claims generally are interpreted as real and proven features of the innovation, the distinction between real and suspected features can be difficult to establish[5]. The *ex post* interpretation of claims in rapidly developing fields, where changes in the dominant approach affects the interpretation of claims, may be challenging despite efforts made to write them clearly at the time of filing.

Patentable subject matter is varied. Examples could include a process or product, a composition of matter (such as a chemical composition) or machine, or a new and useful improvement on any of these. Indeed, as a result of a series of court decisions[6] patentable subject matter in the US has broadened over the past thirty years to include the products of genetic manipulation, software and business methods. Indeed, "anything under the sun made by man" could potentially be patentable according to one decision[7]. Patentable subject matter in the US remains relatively broad compared to other countries, despite extensions that have occurred elsewhere. This partly stems from some differences in the general philosophy of patentable subject matter, tending more towards technicality and industrial applicability in Europe than in the US. These differences have been cited as resulting in the slower movement in Europe towards allowing patents in areas such as business methods, genetic material and surgical methods[8].

---

[3] In other words, an innovation can be patentable but still infringe another patent.
[4] See Guellec and van Pottelsberghe (2007), especially pp. 39-41.
[5] See Guellec and van Pottelsberghe (2007), p. 139 and Bidgoli (2009).
[6] See, for example, *Diamond v. Chakrabarty*, 447 U.S. 303 (1980), *Diamond v. Diehr*, 450 U.S. 175 (1981) and *State Street Bank & Trust Co. v. Signature Financial Group, Inc.* 149 F.3d 1368 (Fed. Cir. 1998).
[7] Contained in *Diamond v. Chakrabarty,* 447 U.S. 303 (1980).
[8] For a discussion of business method patents in practice in Europe, see Harhoff, D., Wagner, S. (2006). For a more general discussion of trends in patentable subject matter in Europe, see Guellec and van Pottelsberghe (2007), especially pp. 119-132.

In most patent systems, a patentable innovation must represent a significant innovative step. Indeed, one of the distinctive features of a patent compared to other intellectual property protection such as copyright is that patent systems have often been thought of as applying to relatively large advances. In other words, the strong exclusionary patent right is granted only in exchange for the disclosure of "valuable" information. In point of fact, the requirement of significance differs across countries and has differed over time. In the US, a patentable innovation is required to be non-obvious as well as new. Judging whether an innovation indeed satisfies the conditions of patentability is a main role of the US Patent and Trademark Office (USPTO). In contrast, some patent systems have traditionally been mere "registration systems", where little or no screening is done to weed out little from big steps. Jaffe and Lerner (2006) document a recent trend in the US towards weaker requirements of "significance" due to workload, incentive system, and other pressures at the USPTO. Hence the concept of significance cannot be considered static but instead responds to intentional or unintentional changes in patent approval practice. The European Patent Office (EPO) has attempted to unify for European states the inventive step that is required in order to qualify for a patent. This requirement has varied across states from a "scintilla of invention" to a relatively high standard of novelty and non-obviousness. In both the US and Europe, the evaluation of significance is generally thought of as an evaluation of how big a step the innovation makes in a scientific or technical sense. Commercial success can be used *ex post* only in a limited way as an argument that an innovation was significant, with the US viewing this sort of evidence somewhat more favourably than Europe[9].

Once granted, a patent may be exercised, traded (sold or "rented" via a licensing contract, or otherwise transferred) or abandoned, like other forms of property. Indeed, contracts of trade are very common, amounting to somewhere between 10 and 20% of the patents issued[10]. While it is common for licenses to be agreed *ex post* -- after a patent has issued -- this is not the only timing that is observed. Licenses can, in principle, be agreed before discovery or even before investment in a research path has begun. These sorts of prospective or *ex ante* agreements specify sharing arrangements for any patents that *might* issue as a result of a research programme. A wide variety of pricing arrangements from no price at all to up-front fixed fees, per unit or revenue-based royalties, profit shares, reciprocal trades of other patents, or other in kind payments can be observed in licensing agreements[11]. Even if quite standard pricing schemes, such as simple royalties, are specified, contracts can vary as to how these payments are spread over time.

Patents differ from many other forms of property in the sense that they are temporary and not permanent rights. Patent protection lasts a statutory maximum of 20 years from the date of filing. In the US, this represents a change from the 17 years that were available from the date of grant prior to the 1994 agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS). Indeed, the patent term has varied over time in a wide number of countries[12]. Statutory protection need not last this long, however, as periodic renewal payments often are required to maintain the right up to its statutory maximum term. Patents frequently are allowed to lapse. Only about 8%

---

[9] See Guellec and van Pottelsberghe (2007), p. 137.
[10] See Guellec and van Pottelsberghe (2007), p. 92.
[11] See Anand and Khanna (2000) for observed license contract structure.
[12] See Jaffe and Lerner (2006), especially pp. 82-94.

of all patents go to full term in Europe, although considerably more reach term in the US[13]. This discrepancy is likely due in part to the differences in the maintenance expenses involved. When translation, maintenance, processing and external (including legal) costs are included, a European patent valid in all member states could cost ten times more than a US patent for a 20 year term[14]. Extension of the patent term beyond 20 years is clearly much more difficult, and can only be attempted by indirect strategies[15]. A "continuation", whereby modifications of an original patent application can be filed over time, can be used to attempt something like a term extension in the US. Continuations constitute a significant amount (about a third) of US examiner workload at the moment, so they are not rare occurrences[16]. Indeed, abuse of this system has been discussed recently by Jaffe and Lerner (2006)[17]. In Europe, the closest proxy of term extensions is probably to exploit so called "dependent" patents. Generally, however, the scope for this sort of behaviour in Europe appears more limited than in the US[18].

Patent Agreements and Administration

The basic system of patents that is used in the United States is set forth in Article 1, Section 8 of the United States Constitution where Congress is granted the power "To Promote the Progress of Science and Useful Arts by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries". Under this power, Congress has enacted various patent laws, the first in 1790, with various revisions added over time. These are codified in title 35 of the United States Code. Congress has also created the United States Patent and Trademark Office to administer these laws and perform other duties related to patent protection. More recently, patent laws in a variety of countries, including the US, have been aligned under the TRIPs (Trade Related Aspects of Intellectual Property) agreements. Other key treaties have included the Paris Convention (which specifies that a first patent filing date in any member state can serve as the patent application date for any subsequent member state filing), the European Patent Convention (which establishes the European Patent Office as a means of coordinating patent grants within Europe) and the Patent Cooperation Treaty (which establishes a uniform procedure for filing patent applications in the member states).

In Europe, patents may issue from either the patent office of individual countries or from the European Patent Office, or both. In fact, a common practice is to file at the EPO only after having filed at the patent office of a specific European country or the

---

[13] See Guellec and van Pottelsberghe (2007) p. 148 for European figures. For the US, the percentage reaching term appears to be about a third, see Lemley (2001).
[14] See Guellec and van Pottelsberghe (2007), ch 7, for an extensive comparison of operations at the USPTO, EPO and Japanese Patent Office (JPO). Under the London Agreement, which came into force in May, 2008, translation fees are anticipated to fall considerably.
[15] A brief scan of the web leads to hits on a variety of ideas to "game the system" to obtain effective extensions. http://www.mewburn.com/Patent/US_Patents:_Term_extensions.htm, accessed 29/7/08 for one such example.
[16] See http://www.uspto.gov for details. For a more complete discussion of strategic uses of continuations, see Lemley and Shapiro (2005).
[17] More generally, strategic (ab)use of filing procedures has been documented in van Zeebroek and van Pottelsberghe (2008) and references therein.
[18] See Guellec and van Pottelsberghe (2007), p. 145.

US[19]. The administrative body responsible for implementing patent grants, such as the USPTO or the EPO, reviews patent applications to determine if the candidate invention satisfies the minimum standards for patentability: novelty and nonobviousness being the most salient characteristics. If the candidate technology is determined not to satisfy this minimum standard, the patent can be denied. A patent can issue after the review process has concluded[20]. This often takes years, with about one out of three patent applications being finally rejected in the US[21]. While the period of exclusivity starts at the date of issuance, or grant, of the patent, the patent disclosure -- the information on the nature of the innovation -- is now published 18 months after the initial filing with the patent office in most cases. This represents a change in the US compared to the older system of being published only at the time of patent grant, which was in force prior to the implementation of the American Inventors Protection Act of 1999.

Enforcement

Patents are only as strong as their enforcement. Enforcement is handled privately in the US (for the most part in civil suits) through the court system. For example, if a patent-holder detects infringement within the jurisdiction of the patent, the patent-holder can sue the violator in court. Infringement suits tend to be extremely expensive[22] and, indeed, can constitute a substantial proportion of total research expenditure[23]. Further, the cost can include a substantial joint loss of wealth rather than a simple transfer from infringer to patent-holder (Bessen and Meurer (2008b)). If the court finds for the patent-holder it may impose an injunction on the violator prohibiting sales of the infringing item and may impose monetary compensation of another type, such as damages. Indeed, a temporary injunction may be imposed even before this. Of course, the defendant can counter-sue as well. A common response to an infringement suit is a counter-claim that the infringed patent was not valid in the first place. Overall, few patents – on average 1.5% of all patents granted -- are ever litigated, and fewer still -- on the order of 0.1% -- ever go to trial[24]. Those that are litigated appear to be the high value patents and those drawn from a subset of

---

[19] This may be done in order to obtain several chances at obtaining a valid national patent. As the EPO tends to produce feedback on the patentability of the innovation slowly compared to national offices or the US, this strategy also has the advantage of providing valuable early feedback to the applicant as to whether the application is worthwhile to pursue. Furthermore, as the USPTO allows more substantial changes to be introduced to the patent document after filing, it may make sense to file at the USPTO early and then file a better-crafted document at the EPO later. See Guellec and van Pottelsberghe (2007), p. 155-159, for patent filing routes.

[20] Other possible outcomes, which differ between the US and Europe, include withdrawal or revision of the document. See Guellec and Van Pottelsberghe (2007) or Jaffe and Lerner (2006) for details.

[21] See http://www.uspto.gov/web/offices/com/speeches/07-46.htm for example. From this press release, it is clear that there is considerable variance across years included in this figure. See also Harhoff and Wagner (2005). Ebert (2004), commenting on the role of patent continuations in the US system, obtains a somewhat lower adjusted rejection rate of ¼.

[22] See Bessen and Meurer (2008a) for estimates.

[23] Lerner (1995) estimates that cases begun in 1991 in the US would eventually total legal expenditures, in 1991 dollars, of about 27% of all of the basic research expenditure in the US in 1991.

[24] See Lanjouw and Schankerman (2001), and Lemley (2001). In other words, about 95% of litigated patents settle out of court.

particularly litigious technology areas[25]. The success rate for patent holders in trials has risen over time[26], a change often attributed to the creation of the unified Court of Appeals of the Federal Circuit in 1982.

The European approach relies much more on an oppositions system to weed out "bad" patent grants relatively promptly. Third parties may submit opinions during the patent examination – to a limited extent -- on whether a patent grant should be made and to a greater extent during a centralised post-grant oppositions procedure[27]. This is a less expensive route to challenge patents in terms of both time and money than full litigation in a court of law and is accessed somewhat more frequently than litigation in the US: somewhat over 6% of all issued patents are challenged in this way[28]. Even though few granted patents go through oppositions, the system and the potential to go through oppositions is cited as a pillar of quality control at the EPO, allowing bad patents – those that do not "in truth" satisfy patenting criteria -- to be screened out shortly after issue[29].

Aside from the oppositions procedure, a second difference in enforcement between Europe and the US is that an EPO patent is, in fact, a bundle of national patents (in the countries designated by the applicant). National laws apply to these patents, with any legal challenges being pursued at the national level. In other words, there is no unified court to deal with patent disputes for the whole of Europe. This can make for multiple litigations in many European jurisdictions with potentially contradictory outcomes[30].

A final area of difference across the Atlantic is what constitutes infringement. A "doctrine of equivalents" may be used to judge whether infringement has occurred even if the infringing item is not a perfect replica of the patented invention. In other words, the looser the interpretation of what constitutes the invention, the greater the "breadth" of the patent protection. Further, an "experimental exemption" allows use of the patented technology for research purposes. Europe has had a more restrictive interpretation than the US on the doctrine of equivalents, but has tended to take a broader view of the research exemption when it comes to university research. This helps explain why infringement decisions for similar types of patents sometimes differ in Europe and the US[31].

---

[25] See Lanjouw and Schankerman (2001, 2004) and Lerner (1995) for statistics and discussion. While some industries have litigation rates as high as 6%, Lanjouw and Schankerman find, in fact, a much lower average rate of litigation at less than 2%. See also Scotchmer (2004), chapter 7 for an overview.
[26] See Jaffe (2000). Allison and Lemley (1998) find that, of the patents litigated to a final determination in their dataset, 46% are held invalid.
[27] Hall et al. (2003) and Harhoff and Reitzig (2004) document the use of this system. See also Guellec and van Pottelsberghe (2007), p. 176.
[28] See Harhoff and Reitzig (2004) for figures and detailed discussion of EPO oppositions. The frequency of opposition seems to have fallen modestly over time.
[29] See Guellec and van Pottelsberghe (2007) p. 178.
[30] The draft European Patent Litigation Agreement aims to centralise litigation, but has not yet been adopted. For more information see http://www.epo.org/patents/law/legislative-initiatives/epla.html.
[31] An important recent decision involving a change in the interpretation of the doctrine of equivalents in the US is *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722 (2002) . See also Weston (1998) for a comparative treatment of the doctrine of equivalents in Europe and the US. EU Directives 2001/82/EC Articles 13(1) to 13(7) and 2001/83/EC Articles 10(1) to 10(5) attempt to harmonize European approaches to the experimental exemption. For discussion of cross-Atlantic differences in the treatment of university research, see also Guellec and van Pottelsberghe (2007) p.

Summary

This brief sketch of the patent right suggests a number of policy levers that can be exercised in order to affect the innovation incentives that actually derive from the patent system. Clearly, one set of policy levers applies to the design of the patent right itself: notably the statutory length of protection, how broadly we interpret the exclusive right, what is disclosed and when this occurs, the size of the inventive step required to earn protection, and what constitutes patentable subject matter. Intervention exercising these levers has occurred in both the US and Europe over time, resulting in a process of evolution of the patent right[32]. A second set of levers applies to the administration of the patent, including how procedures implement the general aims of patent protection. These include the structure of the review process, the structure of fees, the incentives of patent examiners and other administrative features[33]. Again, all these features have evolved over time. A third set of levers applies to the enforcement of the patent in court and includes success rates for patent holders in infringement suits, the types and sizes of remedies imposed and the fora in which patent defence and attack can occur[34]. A final set of levers applies to the freedom with which patent holders can exercise their rights under other bodies of law. A main case in point is the effect of competition policy on the amount of profit that can be extracted from the patent right[35]. This could vary from simple limits on excessive pricing to limits on the ability or necessity to license as well as the licensing contract structure. The literature has examined interventions at all these levels and, in some cases, interactions between different levers. A case in point is the interaction between enforcement methods and patent quality, which will be discussed later. Translating the available administrative and legal policy levers into features of economic models that accurately reflect these levers is, of course, a major challenge. We now turn to how this challenge has been addressed.

III. Economic Interpretation of the Patent Process

We will now use the US system to outline issues in the interpretation of the patent process that determine the basic building blocks on which the rest of the modelling rests. The issues we deal with are the objective function of policy-makers, the source of value of the patent, and the basic functions of the patent that determine its private and social welfare effects. By and large, these basic features are shared with non-US patent systems, so the focus on the US is for expositional convenience.

Interpreting the Goal(s) of Patent Policy

190-191. A recent and significant case on the experimental exemption is *Madey v. Duke University, 307 F.3d 1351 (Fed. Cir. 2002)*, summarised in Janis (2003).

[32] For example, the recent patent term extensions in the US in accord with the TRIPS agreements and the somewhat slower evolution of patentable subject matter in Europe versus the US.

[33] See Jaffe and Lerner (2006) and Guellec and van Pottelsberghe (2007) for a detailed discussion of change in administrative procedures in Europe and the US.

[34] Allison and Lemley (1998) and Jaffe and Lerner (2006) discuss how the move to a unified court of appeals in the US has changed success rates of litigation and litigation strategy.

[35] A simple example t keep in mind for later parts of the chapter could be strict enforcement of EU Article 82(a), which forbids "excessive pricing" by a dominant firm.

Article 1, Section 8 of the US Constitution is quite explicit that the objective of the intellectual property rights system is the progress of "Science and the Useful Arts" [36].. If one were to take this at its word, one would not necessarily want to use social welfare – or even economic growth -- as the standard of optimality in a model of the intellectual property rights system. Instead, one might wish to use the rate of innovation or, less directly, the rate of research and development spending: the more the better. The interpretation one takes is important to the conclusions one reaches about the optimality of any intellectual property protection system. For example, Horowitz and Lai (1996) compare the optimal design of patents when the objective is to maximise the rate of innovation to the optimal design when the objective is to maximise discounted consumers' surplus. A system that aims to maximise consumers' surplus places more value on frequent innovation than a system that maximises the rate of innovation since intermediate steps generate surplus gains for each quality step that enters consumption. Despite the ambiguity in how one should interpret the goal of establishing a system of intellectual property rights in the first place, however, the bulk of the economics literature has taken social welfare to be the appropriate objective that is maximised by policy-makers.

A second issue of interpretation of Article 1, Section 8 is how we understand "*progress*" in "Science and the Useful Arts". Most models capture the significance of patented innovation by some value which is created for society. In some models this value is interpreted as a private market value[37]. The actual patent approval process does not make such a direct link between commercial and scientific or technical value, however. Indeed, the patent document and the review process identify a technical value as well as the source of that value quite explicitly, and also judges "usefulness", but does not go farther to determine any kind of monetary value. Therefore, the patent office makes no direct judgment at the time of grant on market or any other private value to the inventor, has no particular expertise in this area, and does not make market value an explicit criterion for patentability[38]. Protection is not tailored *ex ante* to such a notion of market value.

While the link between value that could be captured by a profit maximising firm and value in terms of technical progress could, in principle, be quite tenuous, as a practical matter they are more closely linked. Since patents are expensive to obtain inventors who are concerned with their own profits would not apply for patents if they anticipated no resulting private commercial value. This leads to the interpretation, taken by Scotchmer (1999) and Cornelli and Schankerman (1999), that inventors have – and seek -- information about whether or not a patent will generate private market value even if the patent office has – and seeks -- little information on this count. The inventors' information is revealed by their patenting behaviour. In particular, inventions with higher private value may precisely be those that are observed to be patented and where that patent is observed to be renewed despite renewal fees.

---

[36] Not all patent systems have a constitutional basis, so in this sense a focus on the US system is somewhat special. While we will not pursue the consequences of this institutional feature of the US patent system, Nard and Morriss (2006) argue that constitutional patent law strengthens the bargain between the state and inventor compared to systems such as patronage. See Jaffe and Lerner (2006) and Scotchmer (2004) for histories of the patent systems and their legal bases.

[37] For example, Gilbert and Shapiro (1990) interpret this value as a flow rate of profits.

[38] As was pointed out earlier, however, the commercial value of an innovation can be brought in a limited way in some jurisdictions as evidence *ex post* that an innovation was a "non-obvious" or "useful" step.

Hence, (continued) patent protection is correlated with innovations that have higher private value in the eyes of their inventors[39].

Finally, regardless of the *magnitude* of the invention's contribution, there is a question of *when* the value is realised for society or for the inventor compared to when the patent right is awarded. La Manna (1994) points out that if the patent right is awarded early, before much of the expenditure to develop the innovation has been incurred, then the exclusionary right ensures that the patent holder can reap the entire reward to its expenditure before that expenditure is incurred. If the right is granted late, and many firms may compete for that right, then the potential patent holder only faces an expected benefit at the time of its research investment. The difference between these two scenarios can affect the incentives to invest, as the patent holder is in a "race" for the right to the fruits of its investment in the latter case but is not in the former case.

More generally, the issue of "when" a stream of social benefits is created also turns on "how" the benefits are created. The social and private value of a patent need not flow directly from the technology that is patented, but may be largely derivative. Value may flow primarily from the innovations a patented advance inspires ("follow-on innovations") or from companion innovations that are used together to create a valuable product ("complementary innovations"). In both these cases, a single patent in isolation may have no private value at all. Indeed, a case we will examine later is that of "pure research tools": innovations that have technical value but no monetary value in isolation. In such cases, a main function of the patent right is to facilitate the transfer of value via licensing contracts from the follow-on innovations or the complementary innovations back to the holder of a "key" patent. Indeed, Hall (2007) presents evidence that patents do actually facilitate such trade in intellectual property.

The Reward and Contract Theories of Patents

Article 1 Section 8 goes on to specify a method that should be used to achieve its stated goal of the "Progress of Science and the Useful Arts". Specifically, a system of exclusive rights *to make, sell and use* the innovation is granted for a limited time. There are two main ways one can imagine that the patent right could promote the progress of science.

The first mechanism is the establishment of a private reward for innovation. This is sometimes called the "reward theory" of patent protection and is presented in the classic work of Nordhaus (1969). The argument is that by generating potential monopoly power – and thus patent monopoly rents – exclusivity provides remuneration for successful innovators. If the cost to generating an innovation is privately borne, then the anticipation of such private compensation is a necessary "reward" to induce innovation in a market setting with profit-maximising agents. If exclusive rights were not available to the innovator, and if the underlying knowledge is a pure public good, any party could use this information to duplicate the invention and compete with the patent-holder to provide it to purchasers. This kind of competition could reduce the rewards to innovation to the point where it would not be worthwhile to conduct the activity in the first place. Hence, the patent system

[39] Many papers have examined patent value and its correlates, see Bessen (2006) and references therein.

promotes innovation that would otherwise be underprovided by the market due to a positive informational externality.

Consider first the "classic case" where a single inventor has exclusive rights to supply an invention that is deemed useful. This inventor is, then, a monopolist over some (residual) demand curve. If the inventor sets a single price, as a monopolist, it can earn profits labelled π in figure 1. This is the private "reward" for the inventive effort. Of course, these monopoly profits come hand in hand with consumer's surplus, s, but also a deadweight loss, d, created by the monopoly pricing. Hence, there is a social cost to ensuring the reward to innovation. The private value captured by the inventor is less than the social value created by the innovation: only by awarding the entire social surplus, the triangle $W = (\pi+s+d)$, could firms' incentives be brought in line with society's. Hence, the incentive to generate scientific progress, while positive, is socially too low in such a system, creating a dynamic welfare loss.

Figure 1 about here

Despite this argument, one cannot conclude, in fact, that a system of temporary exclusion rights necessarily creates socially insufficient incentives to innovate. Exclusivity also generates forces that can create socially excessive incentives to innovate. The patent right designates no single party that is allowed to *attempt* the innovation: anyone can potentially compete for the intellectual property right and the benefits derived from it. In fact, if there are several potential innovators who can compete for the right to earn the "reward" to innovation, there may be socially too strong incentives to invest in innovation. Each potential innovator has the incentive to win the prize to "steal business" from its rival. In other words, the competitors for this prize do not take into account a negative externality that each exerts on others when making an effort to win, leaving losers with nothing to show for their – privately and socially costly -- efforts. Hence, even in the context of a single innovation if there is more than one potential innovator, the market may generate socially excessive incentives to innovate.

More formally, and following Scotchmer (2004), consider a case where two firms are potential researchers. Successful innovation by either firm generates social value W, but there is no additional social value generated by duplicative innovation. Let the prize a firm wins for being the sole innovator be π. Suppose that innovation is probabilistic, so investment generates a probability, p, of successful innovation for the investor and a probability $1-p$ of failure. Of course, it could be possible that both would simultaneously be successful. In this case, the prize is split evenly between the two firms. In this framework, the contribution of each additional researcher to social surplus, when the success probabilities are independent and uncorrelated, is

$$p(1-p)W$$

Having a second researcher is only valuable if the first researcher fails but the second is successful. If this exceeds the incremental cost of research of an additional researcher, then *society* would benefit from the effort of a second researcher. The private contribution, which determines whether a single researcher actually enters or not, is the reward to being the sole winner in the event of failure by the other researcher plus the reward in the event of joint success:

$$p(1-p)\pi + p^2(1/2)\pi = (p-p^2)\pi + p^2\ (\pi/2).$$

If this exceeds the incremental cost of research, then a private firm would benefit from entry.

There are two differences between these two expressions. First, the value is W in the former and only π in the latter. In this sense, there is an under-incentive to conduct research. Now, refer back to figure 1 and set $W=\pi$, so that π reflects the entire social surplus of the innovation. The first term in the private incentive is now the same as the social benefit; however, we also have a second term in the private incentive reflecting the gain to any one inventor when both firms invest and are "lucky". This raises the private incentive for the inventor. An analogous term does not enter into the social benefit because society does not obtain any more surplus when both firms succeed than when one succeeds: society only cares that someone – anyone -- makes the discovery. In contrast, individual inventors care very much who succeeds: the winner obtains a prize whereas the loser obtains nothing. Hence, the private incentive exceeds the social incentive due to this second term. In fact, if the research strategies are perfectly positively correlated (so that success by one firm always accompanies success by the other firm), there is no benefit at all to society of adding a second researcher. On the other hand, there is a positive private benefit for each individual firm since each stands an even chance of receiving the prize. Hence, the patent reward for innovation can create incentives to invent, but these incentives can – in general – either exceed or fall short of the social optimum.

A conclusion from this line of reasoning is that the basic assumptions we make about entry conditions into new research trajectories are likely to affect the conclusions we reach about whether incentives to invent are likely to need to be increased or decreased. If the trajectories are publicly known and the resources to pursue them are freely available, we may be rightly concerned about excessive incentives to innovate. If ideas are revealed to innovators in a unique way, then the concern about socially excessive entry into a "common pool" may be irrelevant. We may need to increase the incentives to invent in this "private information" case.

There is a second mechanism by which the exclusionary rights of the patent can create benefits from invention and "promote the progress of science". When innovations are created, so is information. This can be specific information on the nature of the innovation or it can be in the form of showing that a particular approach to a problem is possible and fruitful. Creating this information is privately costly, but socially useful since it can facilitate innovation by others. The usefulness may be in entirely different fields or markets, so the reward to private innovation need not fall below the investment cost. Still, there is a positive externality exerted by the creators of information. This drives a wedge between the private and social incentives to exert inventive effort, suggesting that information provision needs to be encouraged as it will tend to be under-provided in the market[40]. In particular, even if adequate rewards are available to induce the creation of information, it may be held secret.

[40] Before, $W>\pi$ due to monopoly pricing that caused a deadweight loss. Here, $W>\pi$ even if pricing is not an issue.

The diffusion of information is an explicit goal of the patent system. The "contract theory" of patenting views patents as "contracts" between inventors and society where the patent right is granted in exchange for diffusion of the information that is useful to imitators, followers, or others who exploit the information to generate further social gains[41]. The patent documents create a – freely accessible – library of such information. Even if the rewards to patents and secrecy were the same for the original innovators (so that the same incentive to create information exists) disclosure considerations could make the patent system superior due to the benefits to others. Of course, this argument relies on an assumption that systems that *allow* secrecy actually *generate* secrecy and that the patent library generates accessible and interpretable information. We will consider these issues below.

Summarising, the "reward" and the "contract" theories of patents form the underpinning of how patents generate scientific progress. Reward and information benefits can be seen as underlying both single innovation models of patent design and models of multiple innovations, which we review below but their weighting determines the focus and design of the model. Before discussing these, however, we look at a benchmark where we assume that no intellectual property protection is available. In this framework, we will see under what conditions we can, indeed, generate innovation incentives despite the lack of protection.

IV. Incentives to Innovate in the Absence of Intellectual Property Protection

Is a system granting exclusive rights to innovators necessary to generate a reward or disclosure? Let no intellectual property right exist. Further, as soon as an innovative product is sold or used let a variety of individuals become familiar with the invention, creating the seeds for imitation. If the innovation generates profits, potential imitators are attracted to the innovation to produce their own versions of it. This process creates a variety of suppliers of the innovative product or process, driving down its price and so the profits of the original innovator. If this process is quick or very cheap, then very little surplus is captured by the initial innovator. Indeed, if the cost of developing the innovation in the first place was privately borne, the rapid imitation can reduce the benefits from innovating below the original cost. Any innovator anticipating this process will not invest in the innovation in the first place. In essence, the innovator contributes to a common pool of knowledge when she creates and practices an innovation. This positive externality, if it is not captured by the inventor, generates a private under-incentive to innovate. The patent resolves this problem by making the embodiment of the innovation – in other words, the "object" that is actually traded in the marketplace – a private good even though the underlying knowledge remains a public good.

Many objections have been raised to this argument. These focus on its underlying assumptions that monetary gain is the motivator for invention, and that the imitation process is quick, costless, and purely duplicative. A related point is that the patent

---

[41] See Denicolo and Franzoni (2004) for discussion. Also see Eisenberg (1989) and Miller and Davis (1983) for more complete presentations of legal theories of patent rights linked to diffusion and the "natural rights" doctrine. In an early paper, Arrow(1962) pointed to patents as a way of encouraging information disclosure.

system may be the wrong solution to the problem, to the extent that a problem exists at all. We will address only the former points here, leaving optimal design issues to later in the chapter.

A common concern is that many inventors do not care about monetary gain and are altruistic givers to society who place utility value on the creation of this externality or simply place utility value on the process of creation[42]. Of course, since the patent system is voluntary, the fact that rewards are not necessary to induce the creation and sharing of innovation is not really an objection to the system: inventors are always free to donate their innovations to society. Furthermore, while the pleasure of creation is enough to generate prototypes, it may not be enough to generate the much larger development and commercialisation expenditure necessary to bring the innovation to consumers.

Alternatively, private incentives to invent for individual inventors may derive mainly from signalling to the job market. While some means of attributing innovation to its creator is necessary for this mechanism to work (for example, there may be slow spread of information about the innovation to other job market candidates who could claim authorship), patent protection is only indirectly relevant to innovation incentives in this case. Indeed, depending on the competitive structure of the job market, one could generate either socially insufficient or socially excessive incentives to innovate as individuals compete to develop and showcase their talent. Lerner and Tirole (2002) and Lakhani and Wolf (2005) suggest that Open Source code developers often invent to develop their own skills[43]. If this private benefit exceeds the cost of innovation then the innovation will be provided despite the external benefits it might confer on others – and regardless of intellectual property protection – even though social and private incentives might remain not fully aligned.

While some argue that the patent system is not necessary, others argue that even if it is necessary it is not very effective. Survey evidence of Cohen, Nelson, and Walsh (2000) has indicated that managers do not view patents as very effective at generating direct rewards to innovation. While certain sectors, such as biotechnology and pharmaceuticals, appear to get great benefit from patents, various first mover advantages (such as learning by doing) are credited with generating greater reward to innovation than intellectual property rights in many sectors. If firms rely on other "frictions" such as barriers to entry to generate profits from invention, patents may at best be redundant. On the other hand, Farrell (1995) argues that the "honeymoon" period of patent protection may allow these other potentially long lasting first mover advantages to get going. Hence, patents may contribute more to profits than is acknowledged in the survey results. Still, if frictions and not patents are generating the rewards, then perhaps we should consider weakening or eliminating patents, since the patent system is costly to maintain and may generate few benefits.

An early consideration of the benefits of weakening intellectual property rights in the face of frictions is Cohen and Levinthal (1989), who postulate that imitation is not "free and immediate". Rather, imitation is a skill requiring investment and costly

---

[42] See Middendorf (1981) and Maurer and Scotchmer (2006b). Giuri et al (2006) provide extensive and recent information about inventor behaviour and motivation based on the PATVAL-EU survey.

[43] For a discussion of open source institutions and the possible consequences of patenting for the open source community, see Lerner and Tirole (2005).

effort to "absorb" knowledge from the "common pool"[44]. If *absorption* as a function is costly and can be separated from the function of *contributing* to a common pool, weaker patent protection can have a benefit. A weaker patent makes the common element of knowledge greater and so allows a larger pool. As absorption from the pool is costly, there is a friction – or barrier -- that can allow profits drawn off the larger pool despite weak patent protection. Henkel (2004), in a related model, makes an analogy to a juke box: users each individually and privately contribute to the musical enjoyment of all despite heterogeneous musical tastes and the positive externality they create. In a similar vein, Harhoff et al (2003) suggest that a pool of innovations can contribute as an input to a process of improvement that is fostered by the market but also cannot be fully appropriated by competitors due to "idiosyncrasies". Hence, while each innovation contributes to a public pool, all contributors may be able to draw off benefits that exceed their private contributions. *Weaker* intellectual property rights can spur innovation precisely because of the existence of a common pool and the link between weakness and pool size. In addition, the free riding on research by others that is possible in the common pool also has the benefit of eliminating duplicative research spending.

Bessen and Maskin (2007) systematise the friction-based argument and show that even very small frictions can result in the dominance of a system where all information is available to some (exogenous) degree when compared to -- a particular variant of -- the patent system. Bessen and Maskin's result depends crucially on three elements. First, even if they are small, there are frictions that mean that an inventor's current profit is not reduced to zero immediately by imitation: imitation is costly, time-consuming or both. Second, each pool member is both a contributor to and a benefactor of the pool. In other words, the externality from the innovation runs both ways, from and to an inventor[45]. Third, each firm's private share of a common pool of benefits expands with membership in the pool so that "the more the better". This could be due to the presence of complementarities or network externalities or both. Relatedly, there is a benefit to fragmenting research spending across a large number of firms in that the date of discovery is brought forward under increased fragmentation[46]

Clearly, the assumptions underpinning this model apply much more readily to some industries (perhaps information and communication technology) than others (perhaps pharmaceuticals). Further, the benefit of the "open" system is not compared to an optimised patent system. Indeed, Maurer and Scotchmer (2006b) note that the assumptions on the efficiency of licensing determines the relative desirability of the patent or the open system in their framework[47].

---

[44] See Cassiman and Veugelers (2002) and Bloom, Schankerman and Van Reenen (2007) for discussion and references to the vast spillover literature. Contributions to the common pool are termed "outgoing spillovers" of knowledge and absorption is termed "incoming spillovers".

[45] Belenzon (2006) has presented some evidence that spillovers are "reabsorbed" by the initial inventor in the context of sequential or cumulative innovation in a panel of U.S. firms.

[46] In later models of cumulative innovation, we will see that in some models ideas are assumed to be "scarce". The probability of discovery of an improvement to an innovation may be zero if only the original innovator is present because the original innovator may not have the "idea" for the improvement. Participation of an independent entity is necessary to generate further advance after a first step. The Bessen and Maskin formulation is a less "black and white" version of this approach.

[47] For a further model of the benefits of weak or nonexistent patents that relies on frictions, see Boldrin and Levine (2003).

Costly or slow adoption need not imply the dominance of open systems, however. Some have modelled the adoption process more fully so as to diagnose where precisely the benefits of a patent occur compared to a system of free access. One such benefit to patents could be coordination of the adoption process. A paper that presents a simple version of this is Glachant and Meniere (2008). Suppose that demonstration effects facilitate future adoptions. Hence, early adopters exert a positive externality on later adopters that need not be reflected in early adopter behaviour. On the other hand, an upstream patent holder – as a monopoly holder of the technology – will internalise the externality. Furthermore, an upstream patent holder has the tools to control the incentives for adoption by means of inter-temporal price discrimination. Learning spillovers create two types of inefficiency in their model: first, even if an early adoption is socially desirable, it may not be privately profitable. Second, there is an incentive to delay adoption so as to benefit from the fall in cost. It may be necessary to both "kick off" the process and ensure that excess delay does not occur. Indeed, to the extent that imitation constrains the ability of a patent monopolist to price discriminate over time, imitation can be socially undesirable in this model as it can generate sub-optimal patterns of adoption. Of course, while patents are a way to solve this problem of adoption externalities, other government instruments could well dominate them. A monopolist would not necessarily coordinate the market in the same way as a social planner would. The point is, however, that free access may need to be supplemented with some form of intervention in cases where positive externalities to adoption exist. Earlier papers (Katz(1986), Katz and Shapiro (1986) among others) incorporate these same effects into more complex models that include strategic effects, as well as extensive modelling of the source of the adoption externality (such as a network externality on the demand side).

The papers we have examined so far compare a patenting regime to an open system without rights. Anton and Yao (1994) take the polar opposite case of examining how well an alternative system of *secrecy* can reward innovators. In other words, there are no patents, but it is possible – via an unmodelled legal framework -- to keep "iron-clad" secrets. Is a system of secrecy enough to generate private rewards to innovation? Consider Ms. A, who has an innovation that can be kept secret despite its exploitation (for example, the innovation is a process). Ms. A knows that this is a valuable innovation, but potential buyer Mr. B does not have this information. There is a potential market failure in trading this innovation due to asymmetric information: B does not want to pay for an innovation of little value. If the profit potential of the innovation relies on sale, we have an under-incentive to innovate. Ms. A can attempt to resolve this failure by revealing the innovation to Mr. B. Absent patent protection, however, revealing the information to Mr. B gives him all the knowledge he would need to exploit the innovation without payment to Ms. A. How can Ms. A possibly get any reward for her innovation in this case?

If Ms. A has a secret and reveals it to a purchaser to exploit, she still may be able to gather full profits from this secret because she can threaten – credibly -- to expose the secret to a third party, Ms. C. Hence, revelation of the value of the innovation also reveals the magnitude of loss under "punishment": the more valuable the innovation, the more value could be lost if the secret were revealed to Ms. C. More precisely, since the innovation is a secret that is not patented, the threat to reveal comes along with it the *credible* threat not to attach strings to the innovation such that Ms. C. would be a restrained competitor. It is precisely this "weakness" of secrecy that makes it a strong negotiating tool. Further, not only does a threat exist (because Ms. C is available outside the private bargaining between Ms. A and Mr. B) but the strength of the threat is linked to the value of the technology. Now, Ms. A would always approach Ms. C if the payments were not contingent on profits actually earned, but if the license payments are contingent on the gross profits actually earned by the licensees, which are assumed to be observable and verifiable, then Ms. A could prefer to maintain a monopoly structure as long as she gets a share of the gains. Under the relatively mild assumption that the gains from maintaining a monopoly structure rather than duopoly are large while the profit gains for Ms. C from accessing the same technology as Mr. B are small, secrecy can be associated with trade and also a reward to innovators[48].

This argument relies on the parties' being able to keep a secret so that imitation cannot simply destroy the innovation's value out of the control of the parties to the transaction. Either one can interpret this as an implicit friction in the model that allows for profits to be reaped despite the absence of strong property rights or one can interpret this as a very robust legal framework enforcing an -- alternative -- trade secrecy system. Clearly, the situation analysed is quite special in that it relies on the innovator's not being able to exploit the innovation herself after revelation, as this could affect the negotiations. Nor can the innovation be reverse engineered based on *ex post* observation of market products. Hence, secrecy comes with a great deal of control in this model, in contrast to the story we told at the beginning of this section where secrecy really was not an option. Their paper suggests, however, that in situations where information can be controlled extensively sufficient rewards to induce innovation can be created even in the absence of any *patenting*. Furthermore, a complete lack of *patent* rights is consistent with some disclosure when this degree of control exists. When we compare a patent system to an alternative of secrecy, we should not automatically assume that all information is disclosed under patenting and none under secrecy.

Hence, one can make a theoretical argument against patent rights if the circumstances are right. If frictions make patents redundant, or if secrecy is a very effective tool, then they may not be worthwhile. Do the conditions exist for weaker protection systems to generate more innovation? Empirically, this is not yet a settled question. Hall (2007) suggests that strengthening a patent system (in terms of lengthening the patent term, broadening subject matter coverage or improving enforcement), while associated empirically with more use of the patent system, has less clear effect on aggregate innovative activity. Indeed, Sakakibara and Branstetter (2001) found that the effect of strengthening patents in Japan had only a very small effect on R&D

---

[48] Formally, for Mr. B to accept the license it must be the case that the (contingent) fee not exceed his gains from moving from symmetric duopoly to monopoly, $\pi_M - \pi_D$. For Ms. A not to approach Ms. C, it must be the case that the fee from B exceed what she could earn from Ms. C ($\pi_D - \pi_L$), where π_L is the profit Ms. C would earn if excluded from the innovation, plus the fee that B would pay knowing that the secret would be disclosed afterwards to Ms. C. If $\pi_M + \pi_L > 2\pi_D$, as is true for many standard industry structures, then a "wedge" exists such that even with a positive fee in the case of disclosure to all parties, a parameter range exists where disclosure of a valuable innovation to only Mr. B is the equilibrium.

activity[49]. In a cross-country study using a broad measure of patent strength, Lerner (2005) finds that strengthening patents has a positive effect on innovation if patents are very weak, but a negative effect if patent protection is quite strong, so that intermediate levels of protection seem to work best at inducing innovation. The test is a step away from the specifics of the theories discussed here, however, so it is difficult to tell what mechanism is at play. To the extent that there is a linkage between innovation and patent strength, the main effect appears centred on a few industries where patents tend to be viewed as quite effective, notably pharmaceuticals, medical devices and biotechnology (see Arora, Ceccagnoli and Cohen (2003)). In section VI, below, we will consider optimal alternative systems of rewarding innovative effort that, combined with free access to innovations, might dominate patent systems.

V. Optimal Patent Design: Length and Breadth of Protection

We have argued that patents can be seen as generating some kind of social benefits by rewarding technical progress, but it may be hard to argue that they generate progress optimally due to the deadweight loss they potentially create. We have also said that there are reasons to believe that the reward of a patent could generate either socially insufficient or excessive incentives to innovate —in addition to generating the deadweight loss. The argument for moving to no intellectual property protection at all or to a system of pure trade secrecy relies on relatively specific assumptions, however. An alternative would be, then, to retain the basic features of the current system but re-balance its parameters to generate improved performance. There is a long tradition of papers within this approach. We will consider first models where there is only a single innovation, followed by models of multiple innovations. These models will be primarily in the "reward theory" style. We will look at models that focus on disclosure issues in the final section.

Single Innovation Models

A first set of papers examines a single innovation. Nordhaus (1969) set the stage for this work by suggesting that the length of patent protection should balance off two forces: first, for an innovation that will potentially yield benefits to society forever, the length of protection should be long (potentially infinite). Since protection is based on exclusive ownership, however, this creates a potential deadweight loss due to monopoly pricing. Minimising this deadweight loss argues for short protection. The optimal length of protection needs to balance these concerns: the longer the protection the more innovation is induced, but the worse is the deadweight loss. Suppose that the innovation generates a notional maximum discounted social value $\overline{W}$ that could be earned if it were available for free immediately but a deadweight loss, *d*, is incurred during each period of protection[50]. Let the flow profits for each period of protection be π for the innovator. Profits fall to a baseline level of zero after protection expires.

---

[49]Empirical investigation raises the issue of simultaneity between research intensity and intellectual property protection. Studies controlling for this link (see Qian (2007)), however, continue to find relatively little evidence of a relation between strength of protection and investment activity.

[50] In the earlier notation, if welfare equals d+s+π per period, as before, society obtains only s+π during the period of protection, but the entire welfare triangle after the patent expires.

If we define $X(T)=\frac{1-e^{-rT}}{r}$ when there are *T* periods of patent protection, the innovation generates net benefits $\overline{W}-dX(T)$. This expression is decreasing in *T*. We can either think of maximising this expression with respect to *T* or minimising *dX(T)* with respect to *T,* subject to the constraint that the discounted benefits generated from innovation, *X(T)π*, meet a value, *c*, required to induce innovation. Noting that *X(T)π* is increasing in *T*, the solution to this problem is the minimum *T* that allows the constraint to be met.

Of course, any tool affecting monopoly pricing, including competition policy, could have a similar effect. For example, one could invoke a limit on prices (via, for example, Article 82(a) in Europe) in each period but allow protection to persist for a very long time so that the small return in each period cumulates to the desired reward. Hence, the recommendation is that patents should last a long time, but should be combined with a strict limit on pricing. This argument was developed by Tandon (1982), where the limit was imposed via compulsory licensing guidelines. Formally, society's problem is to maximise total discounted social welfare from innovating (or minimise the total discounted social costs), where welfare in each period decreases with the price premium over marginal cost and the patent expires at time *T*. If we take *π* to reflect the price cost margin, we have the following social planner's problem:

$$\underset{\pi,T}{Max}\ \overline{W}-X(T)d(\pi)\ \text{(equivalently}\ \underset{\pi,T}{Min}\ X(T)d(\pi))$$

subject to the constraint that

$$c \leq X(T)\pi$$

where *c* is the value that must be covered to induce innovation. The difference between this formulation and the one above is that there are now two instruments of control (the price-cost margin and statutory length) and one, the price-cost margin, is an argument in deadweight loss so that we now have function *d(π)*. Tandon shows for the case of linear demand that while the minimand varies proportionately with the discount factor *X(T),* it is proportionate to the square of the royalty rate (which determines the price premium) via deadweight loss. This makes the length of the patent, *T*, a relatively efficient instrument to compensate innovators relative to the price-cost margin. Commenting on this problem, Ayres and Klemperer (1999) point out that lengthening the patent life so as to hold the expected profit constant while restricting the monopoly distortion is a form of Ramsey pricing: price is set to minimise its distortionary effect while still generating a target amount of revenue[51].

In this model the patent designer wields a great deal of control: both the price-cost margin and the length of time during which that margin can be charged are direct instruments. No imitation limits the time during which rents can be earned (so that the effective patent length is the statutory patent length) and no imitation or other

---

[51] Allowing a supra-competitive price is like allowing the patent holder to impose a tax. It could be even more efficient to spread a centrally-collected tax over all *goods* as well as all time periods, but this would require intervention by a government body that knows the appropriate target level of reward. As we discuss below, this is an unrealistic informational assumption in many cases. See Ayres and Klemperer (1999) for more discussion.

competitive concerns limit the price-cost margins that can be charged. We shall see later that imitation can modify the model results and the policy implications.

Gilbert and Shapiro (1990) extend this line of research in two directions. First, they associate the freedom to charge a large price-cost margin with "patent breadth". Patent breadth can be thought of as a strength index for the patent where a stronger patent is associated with higher flow profits. One interpretation of this would be that broader claims could be approved by the patent office, resulting in a larger "exclusion zone" around an innovation in product space. This could translate into higher monopoly profits if close substitutes are not permissible. Hence, patent policy now consists of two policy instruments: patent length and breadth, where the policy behind breadth is interpreted more generally than in Tandon's work.

Second, Gilbert and Shapiro show that either long, narrow patents or short, broad patents can minimise the deadweight loss cost of patent policy, subject to the constraint that the innovator earns a reward that induces some desired level of investment. Which of these designs is better depends on how welfare is related to profits: the relation is assumed negative, but the second derivative can realistically take either sign. It is this second derivative that determines optimal policy. The welfare maximisation problem they consider is the same as above. Their approach is to define the "required profit", $\pi(T)$, as the value of flow profit, π, that satisfies the constraint for some given level of c. Total welfare, $W(T,\pi(T))$ can then be obtained solely as a function of T. They analyse the optimal policy by considering the shape of this function, W, as T changes. This shape is determined by both direct and indirect effects:

$$\frac{dW}{dT} = \frac{\partial W}{\partial T} + \frac{\partial W}{\partial \pi}\frac{d\pi}{dT}$$

The first term on the right hand side, the direct effect of lengthening the patent, is clearly negative due to the deadweight loss per period of protection. The second partial derivative on the right hand side is the direct effect of increasing profits on welfare. This is also negative via the deadweight loss. The third term, the effect of lengthening the protection on the "required" profit level, is also negative, however. As a result, the sign of the right hand side depends on the weighting of the effects, which is determined by the second derivative of W. In order for an infinitely-lived patent to be the optimal design, we need the entire expression on the right hand side to be positive. This occurs if welfare is concave in the patent-holder's profits, $\frac{\partial^2 W}{\partial \pi^2} < 0$, since the patent reward becomes increasingly costly to welfare via the deadweight loss in this case. The reward should be constructed of flow profits that are very small, but cumulate to the target reward level that we wish to achieve in order to maximise dynamic benefits. Hence, the welfare maximising policy is narrow, long patents. This is essentially the situation analysed by Tandon. Gilbert and Shapiro show that there is, however, a second case where deadweight loss is decreasingly costly as breadth rises (the second derivative of welfare takes the opposite sign). Here, short, broad patents are optimal. The optimal patent design is no longer so clear.

It may not be appropriate to consider patent breadth to be an absolute exclusion zone, however. After all, patent protection does not prevent competition from other firms if they come up with non-infringing substitutes. Gallini (1992) suggests that a broader patent should instead be thought of as one that is more costly -- not impossible -- to invent around, noting that neither the Gilbert and Shapiro nor the Tandon approach really accounted for imitation possibilities. In this optic, patent policy consists of length and breadth – the latter now defined as the cost of imitation. The constraint we have included before[52] is now accompanied by a second consideration that entry will occur freely up to the point where the entry cost, *E*, is just offset by the gains from competing as an oligopolist in the industry. The policy maker must take into account that the increased reward earned by patent holders makes entry more attractive. Specifically, for a given cost of imitation, lengthening the period of patent protection now makes imitation more attractive. The entry costs incurred are, of course, also social costs. Hence, the social planner's problem is to create a reward that compensates innovators, minimises deadweight loss, and minimises duplicative spending. More formally, let the profits earned by any firm actually entering the industry, π, be a continuous and decreasing function of the number of entrants, *m*. We have entry determined by the condition: $E = X(T)\pi(m)$. If we take $m(\pi)$ to be the inverse of this profit function, we have the social planner's problem:

$$\underset{\pi,T}{Max}\ \overline{W} - X(T)[d(\pi)] - m(\pi)E$$

subject to: $E = X(T)\pi$
where we also have $c \leq X(T)\pi$

In other words, we now maximise the surplus net of deadweight loss and entry costs, subject to a free entry constraint, and generating rewards sufficient to induce patenting. The patent length that maximises social surplus when *E* is small enough that the threat of imitation is real will be the one that just discourages imitation -- under the condition that the output elasticity with respect to the total number of firms, $(m+1)$, does not exceed 1. The elasticity condition means that the relative cost of excess entry is large compared to the deadweight loss benefit of an additional entrant. If imitation is just discouraged, there is no resource cost to imitation for any given level of industry profits. Also, any given level of industry profits can be channelled entirely to the innovator if no imitation occurs. Hence, the optimal length for any given *E* is capped such that the policy just discourages all entry ($m=0$)[53]. If both the length of protection, *T*, and the breadth of protection, *E*, are policy levers of the government, then the best policy (when the same output elasticity condition holds) is to set *E* – the direct instrument to control imitation -- large enough to discourage all imitation and the length, *T*, to generate the desired reward for innovation. Hence, we solve for length from the second constraint (assuming no entry will occur), then set the entry cost level such that the first constraint generates no entry. Entry cost, as it

[52] The "innovation constraint" in Gallini's model has a different interpretation. It is a constraint that the innovator uses the patent system rather than secrecy. Hence, while written the same way in our notation, the interpretation of *c* is "the value that could be earned by not patenting" and the constraint is now that patenting as a protection option dominates not patenting and using secrecy. This interpretation is in keeping with the emphasis on imitation in the model.

[53] This can be obtained from the entry constraint.

both loosens the cap on profits earned by the patent holder and also discourages imitators so that all of those profits are channelled to the innovator, is a very efficient instrument in this framework, so there is an argument for patents that are optimally broad and short.

Klemperer (1990) conceives of patent breadth as a zone of exclusion in product space around any given invention: the best current design is protected by patent so imitators only offer inferior products. Hence, increased breadth has the cost of redirecting "imitation" away from desirable designs and into less desirable products, inducing a welfare loss for any consumer who buys a "knock-off" rather than a more desirable product. In this sense, the formulation extends imitation considerations to envisage the possibility of (inferior) imitation and a zone of exclusion. Breadth is interpreted as a portion of a product spectrum reserved for a patent holder[54] with imitation supplied competitively (so that the best non-infringing design is supplied "at cost"). Now, the social cost includes not only the deadweight loss due to a price distortion on the patented product, but also the transportation costs for consumers who "travel" to the patent boundary to purchase there, and any additional reduction in demand for travelling customers due to the cost of transport. Given that imitation is redirected so as to occur outside the zone of exclusion, the pattern of consumption induced by breadth as a tool may be undesirable for society as a whole since it may result in consumption switched to the "wrong" product variety. Call losses from "travel" $\tau(z)$. We now think of the social planner's problem as minimising total discounted social costs:

$$\underset{z,T}{Min}\; X(T)[d(z)+\tau(z)]$$ subject to the constraint that

$$c = X(T)\pi(z)$$

where z is the width of the exclusion zone of the patent, and both deadweight loss and profits, π, vary with this width. Combining this constraint into the minimand, we can reformulate this problem as minimising the ratio of patenting's social cost per unit of money spent with the patent's lifetime set to be the minimum that satisfies the innovation constraint:

Choose z to satisfy: $\underset{z}{Min}\; \frac{c[d(z)+\tau(z)]}{\pi(z)}$ and choose T such that $X(T)=\frac{c}{\pi(z)}$

For very wide patent breadths, the social cost is primarily a simple monopoly distortion since little substitution actually occurs. For very narrow patent breadths, however, there is high competitive pressure from close substitutes so that deadweight loss is low. On the other hand, substitution occurs readily for narrow breadth so that the cost of travel contributes significantly to social cost. Optimal patent design is, therefore, very sensitive to the pattern of consumer preferences. If all consumers have

[54] As patents are specified in *technical* terms, there is no precise legal equivalent of such an exclusion zone in *product* space for many patents. Still, one can think of some patents that are quite close to this specification. For example, a patented shoe (US Patent number 5255452) used in stage illusions includes claims on both lace-up and strap-on variants. To the extent that a "knock off" would have to use an attachment mechanism other than the standard ways a shoe attaches to a foot, this could create the sort of inferior knock-off in product space that this model envisages.

identical transport costs, the patent holder *must* set prices low enough that no consumers switch. If prices are set such that no switching occurs, however, the only social cost that is relevant to the social planner's decision is the standard deadweight loss. As we have just seen, this is minimised by setting the breadth as narrowly as possible. In contrast, if all consumers have identical reservation prices for their most preferred variety (so demand is inelastic), then the patent holder optimally charges consumers this (common) reservation price. All consumers purchase, but there is zero deadweight loss. If the patent breadth is set as wide as possible, then all industry profits accrue to the patent holder and travel costs are minimized. Hence, wide breadth minimises the social cost per unit of monetary incentive for the patent holder. In all cases, the patent lifetime is set so as to just satisfy the constraint that the patent holder have the incentive to create the innovation in the first place: narrow patents must be long; broad patents must be short[55].

These three papers, Gilbert-Shapiro (1990), Klemperer (1990) and Gallini (1992), taken jointly suggest that there is no clear-cut answer to whether larger or smaller breadth is better for social welfare. Gilbert-Shapiro's result that narrow, long patents are associated with a decreasing and concave flow social welfare function is intriguing, but they show that this is not the only case. Indeed, Gallini and Klemperer provide examples where social welfare can be convex or simply increasing in patent breadth. These different shapes give rise to drastically different optimal policies.

Two general comments have been made about this type of patent design story. First, most of these models take the identity of the original innovator as given. Denicolo (1996) notes that reduced breadth may be accompanied by more entry by researchers. If more entry into the research stage is accompanied by the presence of inefficient producers, insufficient product variety, or duplicative research costs, then clearly whether or not increased breadth is desirable on balance will depend on how these costs and benefits weigh up in social welfare. His paper generalises the reasoning of the previous models to obtain sufficient conditions under which maximum or minimum patent breadth is optimal when the identity of the original innovator is determined by means of a patent race. In his formulation, then, we have:

$$\underset{k,T}{Max}\ \overline{W} - X(T)d(k)$$

subject to $c \leq X(T)I(k)$

where *I*, the "incentive" to win, is now a the average of the profits a racer could lose weighted by the probability of losing, and a measure of the profit gain from patent protection, weighted by the probability of winning. The term *I* is a function of the flow profits due to the patent holder. Denicolo postulates that narrower breadth is associated with a smaller profit for winners compared to losers. Hence, denoting a

[55] Waterson (1990) also takes a spatial exclusion zone interpretation of patent breadth. His result is that flow welfare can increase in patent breadth due to better locational decisions by firms. In other words, if an imitative entrant unconstrained by patent breadth would choose to locate socially too close to the incumbent firm for business stealing reasons, an exclusion zone around the first entrant can benefit both innovators and society. For the innovator, an exclusion zone ensures a greater degree of product differentiation, hence attenuating competition, and for society it ensures greater variety, reducing "transport" cost losses.

patent breadth index by *k*, we have the incentive to win and deadweight loss both as positive functions of breadth. Narrower breadth tends to bring about more competition, but what this means in terms of social welfare is ambiguous, since social welfare could decrease (due to duplicative research expenditure in the R&D race or inefficient production) or increase (due to reduced deadweight loss) and could have almost any second derivative with respect to breadth. The paper takes a reduced form approach to this, showing that the optimal patent policy of long, narrow patents or short, broad patents depends on the second derivative of flow welfare with respect to breadth as before, however these second derivatives are obtained. This paper shows, then, that the general reasoning of the earlier models is not affected by the incorporation of racing concerns even if the precise breadth and length cutoffs might be.

Second, Gallini (1984) has pointed out in earlier work that licensing activity has the significant benefit of allowing market participants to economise on duplicative research expenditure. Surely, an incumbent facing a threat of entry by imitative spending should have an incentive to contract *ex ante* with the potential entrant to save on the imitative expenditure. Both parties are at least as well off under this scenario. This sort of insight should suggest that introducing licensing along with imitation concerns could dramatically change the conclusions we draw on the length-breadth trade-off. The precise effect depends on the modelling. Consider the Klemperer (1990) and the Gallini (1992) formulations. In the Klemperer model the entry cost is zero, so licensing activity would not alter the basic conclusions: the patent-holder would potentially have to license an infinity of firms in order to avoid imitation, and this it cannot do. On the other hand, in the Gallini paper, the entry fee is both positive and central to the analysis. *Ex ante* licensing by the patent-holder allows the entry fee of imitators to count as income for the innovator. Hence, it becomes an innovation incentive for the patent holder, but does not enter as a welfare loss because it becomes a pure transfer between market participants[56]. This alters the problem considerably, so that the set-up really becomes one of minimising deadweight loss subject to an innovation constraint that now includes licensing revenue from would-be imitators. This can weaken the argument for broad protection. Hence, when we include licensing considerations, there is a stronger argument in this model for long, low breadth patents.

Cumulative Innovation Models

All the above papers were set in the context of a single innovation, albeit with possible imitators or knock-offs. This is a simple and instructive case, but not necessarily a commonly observed one. Two cases of multiple innovation streams which raise issues quite distinct from those discussed above have been analysed in the literature in response to this concern. The first of these cases is discussed in this sub-section. It is the case of cumulative innovation: where innovations build on previous advances. This case raises new challenges to patent design in the following sense. Suppose that, without a first innovation, the idea for an improvement cannot exist. The fact that the first innovation creates the seeds for its own improvement means that

---

[56] Indeed, in the context of arguing for an independent innovator defence, Maurer and Scotchmer (2002) argue that imitation costs should not determine patent design, as their impact should be minimised by privately organised licensing activity. We will see damages used to (optimally) compensate innovators for infringing in later models in this survey.

there is a positive externality running from the first innovation to the second. This externality need not be internalised if the follow-on innovator is distinct from the first innovator. How to best divide a single profit stream so as to both reward the first innovator for this externality and induce follow-on innovation is the focus of this literature.

More precisely, following Scotchmer (1991, 2004) and using our earlier notation, suppose that one firm has generated an innovation that could give rise to further innovations: without the first innovation, the follow-on would not be possible. For example, think of a basic innovation that opens an entirely new field of research that had heretofore not been contemplated. The first innovation generates a positive externality by its revelation, as it identifies the new field. From society's perspective, the full benefit of the first innovation includes creating the possibility of a stream of innovations that cumulate to produce benefits, ultimately for consumers. If these innovations are separately held by independent inventors, however, we face the challenge of simultaneously generating full incentives for the first innovator to "kick off" the innovation path and also generating full incentives for any subsequent improver to produce follow-on innovations. Suppose, for example, that the second innovation generates a positive total discounted social value of W_2 on its own. If we award W_2 entirely to the second innovator, then we create full incentives to invest in the follow-on. However, we still face the difficulty that we should attribute both the direct value of the first innovation, W_1, as well as the value of the second innovation (which would not have existed without the creation of the first innovation), W_2, to the *first* innovation. Hence, to maintain full incentives to create the first innovation we need to allocate W_2 twice. Otherwise, innovation incentives will be socially too low.

Establishing exclusive rights can partially address this "double allocation" problem. If a single innovator has control over the rights to an entire stream of innovations there is no need to allocate W_2 twice. A social planner or any other single inventor would internalise the externality and so we would not have any trouble achieving efficiency. This solution is straightforward if the same innovator is able to efficiently obtain both the initial innovation and its follow-ons. However, if a single entity does not have the ability to create all inventions that stem from the information revealed by a single invention, property rights may be used to allocate the benefits of the externality so as to achieve the desired technical progress despite the participation of multiple parties. This is where the possibility of licensing matters. Licensing makes it possible for a first innovator who has exclusionary rights to follow-on innovations but not the ability to develop them to trade access to those rights for a benefit flow from the second innovation. As long as enough benefit is left to the second innovator to cover the costs of creating the second invention, it is in the interests of both innovators to agree access and also to conduct research to generate the second invention. In this way, the presence of exclusive property rights does nothing to impede the pace of innovation. To the contrary, property rights facilitate net benefit transfers from future innovations to the first innovator, improving the incentive to develop innovations in the first place.

Using the terminology of O'Donoghue, Scotchmer and Thisse (1998), we must now distinguish in our patent design problem between patent breadth as protection from pure imitation ("lagging breadth"), and patent breadth as protection from different – and perhaps better quality – follow-on innovations ("leading breadth"). In the single innovation case, only the former was relevant. With cumulative innovation, however, the leading breadth granted to the first innovator determines whether a follow-on innovation infringes the original patent and therefore can be barred from sale by the first innovator. A patentable innovation outside the scope of (leading) protection is non-infringing, while one inside this scope infringes. A broad patent on the first innovation implies, then, that a follow-on innovator would need the express agreement of the first innovator to exploit the follow-on. Hence, the first innovator can use the agreement to allocate the externality back to the first in the chain of innovations.

Green and Scotchmer (1995) enunciate and formalise this basic insight. Consider the case of a "research tool" where the value of the first innovation on its own is nil[57]. If there is only one potential innovator, then the fact that there is an externality running from innovation 1 to innovation 2 creates no inefficiency in the decision to innovate as long as the patent runs long enough that the total profit of the innovator exceeds the total development cost. While the innovation constraint becomes $c_1+c_2 \leq X(T)\pi_2$, the problem is essentially unchanged from the single innovation case. The innovation stream should be undertaken whenever $W_2 - c_1 - c_2 > 0$ and policy must set patent length to allow the monopoly profits to cover the full cost of investment, c_1+c_2. Breadth has no new role in this story. If the inventor of the second innovation is different from the first inventor, however, then we must be concerned not only with the total profit but also its division, as both inventors must obtain a large enough percent of the earnings to cover the development cost. Now we have two innovation constraints: the first innovator's earnings over the patent period must cover investment cost c_1, and the second innovator's earnings must cover investment cost c_2. Because the earnings may now include transfers between the parties, bargaining is now the focus of the analysis.

Leading patent breadth can both affect the bargaining positions of the parties to the technology access agreements (licenses) and the need of each party to "come to the table" in the first place. As the first innovation has zero value on its own, the first innovator would never invest unless some of the second generation profit was transferred to it. Such a transfer can occur via a licensing contract, but the timing of this agreement matters: whether this contract is executed before (*ex ante*) or after *(ex post*) the second innovator invests c_2 can determine what terms will actually result from bargaining between the two parties[58]. This is because *ex post* agreement allows the second innovator to be "held up" for (sunk) cost c_2. Hence, the bargaining equilibrium potentially depends both on the breadth of the patent and on the licensing regime that is permitted.

Suppose that only *ex post* licensing is permitted and while both innovations are patentable, the patent breadth is such that the second innovation infringes the first. In this case, either innovator could potentially prevent the second innovation from coming to market. If the firms fail to agree a license, no transfer is possible and the

---

[57] There is some evidence that research tool patents have increased over the last twenty years, in particular in the area of biotechnology, making this a pertinent example. See Walsh, Arora and Cohen (2003).

[58] One could also think of very early agreements, before innovation 1 is created. This would be closer to a research joint venture, on which there is a considerable literature. See Tao and Wu (1997) or references included in Miyagiwa (2007).

first innovator cannot benefit from the second innovation. Further, the second innovator stands to lose its development cost but obtain no return for it if the innovation is blocked by the first patent holder. On the other hand, if the firms agree to split the surplus evenly, the first innovator potentially earns half the profits from the follow-on innovation, as does the second innovator. Hence, *ex post* licensing results in profits ($\frac{1}{2}\Pi_2 - c_1$, $\frac{1}{2}\Pi_2 - c_2$) for the first and second innovator, respectively, and if π_2 is a per period reward, the cumulative reward for a patent of length T is $\Pi_2 = X(T)\pi_2$. is the discounted profit from the follow-on innovation[59]. Because the second innovator earns only half the profits from innovation 2, innovation may be deterred when [$\frac{1}{2}\Pi_2 < c_2 < \Pi_2$] -- due to "hold-up". Hence, *ex post* licensing does not fully resolve the reward problem at both the levels of innovation one and innovation two. Furthermore, if we narrowed patent breadth so that the second innovation no longer infringed the first, innovation 1 would never be undertaken at all: the first innovator would have no basis on which to capture value in exchange for its cost of investment, c_1.

An *ex ante* agreement can resolve this problem by allowing the first innovator to commit to a lower licensing fee by means of negotiating at a time when the second innovator has yet to sink development cost c_2. The second firm can ensure itself a payoff that covers its investment cost. Further, both innovations will obtain enough surplus to be innovated as long as profits cover the entire costs, $\Pi_2 > c_1 + c_2$. Hence, the combination of *ex ante* licensing and large patent breadth for the first innovation generates desirable investment incentives. If the second innovator knows Π_2 and c_2 before investing, the optimal policy is infinite breadth, in fact, so that *all* follow-on products infringe the basic innovation. This minimizes the second innovator's profit in an *ex ante* agreement. In other words, the situation where the second innovator's product infringes puts it in the weakest bargaining position, allowing the first innovator to give it only just enough to induce it to innovate. This "outsourcing" in turn ensures that profits are channelled to the first innovation as a reward for the externality it generates.

The authors comment that the legal status of *ex ante* licensing agreements such as the one we've just discussed is questionable under competition policy since one could claim that they could amount to *ex ante* collusion. On the other hand, if one restricted all licensing to be *ex post*, one would have to recognise that this could restrict the cases where the follow-on innovations are developed or could require that patents be lengthened in order to increase the reward of the patent holders sufficiently to satisfy their innovation constraints. In other words, we would need to "scale up" the term Π_2. Whether or not this is desirable depends on the deadweight loss associated with the patent period. Notice that, since the effect of the stringent infringement standard and *ex ante* licensing is to obtain a better distribution of licensing revenues – which are a pure transfer – the change in "breadth" has no direct effect on deadweight loss. Any deadweight loss is via the patent term, and this can be minimised while inducing both innovations to occur when the first patent holder is given broad control and *ex ante* licensing ability.

[59] The underlying story could be that Nash bargaining determines an even split in the licensing negotiations.

The emphasis of this argument is on the importance of infringement as the salient aspect of "breadth" in a cumulative setting. This, however, is not the only policy tool that could be relevant to the allocation of surplus between initial and follow-on innovators. Scotchmer (1996) investigates how the division of profit is affected by the *patentability* of the second product. This plays a role when the identity of the second innovator is not known *ex ante.*

Again, consider the case when the first innovation has no stand-alone value and assume that the follow-on innovator need not be the same as the first innovator. Let the second innovation infringe the first. The first innovator would potentially issue an exclusive license (before research into the follow-on has occurred) to a single agent, hoping to collect profits so as to provide a payoff to its own basic innovation. If the second innovation is separately patentable, however, then *any* independent second innovator (regardless of whether it was a licensee) can block the follow-on's sale. Suppose, then, that two independent innovators potentially could invest c_2, each innovator potentially patenting the follow-on with probability ½. The first and second innovator must then bargain *ex post* over surplus Π_2 regardless of whether the second innovator had previously received a license. Assume the any bargaining parties split this surplus evenly so that each receives $\Pi_2/2$. A potential second innovator who obtains a license faces a probability of ½ of obtaining the follow-on and earning Π_2 and a chance of ½ of losing. In the latter case, the license to the first innovation is useless without a license to the second. Hence, the licensee ends up with expected profit $½\Pi_2 + ½\Pi_2/2-c_2$. The entire stream of innovation is now expected to generate $3\Pi_2/4 - c_2$ for the licensee. The firm that was a non-licensee earns nothing if she loses the race for the follow-on, but can bargain for $\Pi_2/2$ if she wins, which she does with probability 1/2. Hence, a non-licensee expects to earn $\Pi_2/4-c_2$. The winning bid for the exclusive license, the difference between these, is only $\Pi_2/2$ which is less than $\Pi_2 - c_2$ whenever the losing bid would be positive. If, on the other hand, the second innovation is not separately patentable, an independent follow-on innovator has no exclusionary rights. As a result, the first innovator never bargains with a non-licensee, and non-licensees never invests. The first innovator can earn the entire net stream of returns in this case, $\Pi_2 - c_2$, as a payment for an exclusive (*ex ante*) license.

We see, then, that the first innovator receives a lower payoff when the second generation innovation is patentable than when it is not. Patentability of the second generation product has two drawbacks in this story: it potentially encourages duplicative R&D costs for the follow-on product – reducing the surplus available to the bargaining parties -- and also it transfers some of the profit stream towards the follow-on inventor. We have an argument based on these two papers for very strong rights to seminal innovations but relatively weak protection for any follow-ons.

Of course, policy-makers would generally have more instruments than patentability to work with. As before, we could consider the value generated by the patent, Π, to be an increasing function of patent term, T. The length of the patent serves to scale the reward. In this case, we could examine how patentability and patent term could work together to create rewards for the innovators. If there is no deadweight loss to the patent, then this yields an answer that infinite protection is optimal. More generally, let there be a deadweight loss to protection that we wish to minimise, subject to an innovation constraint. Then to cover the first innovator's cost c_1, patent life could now be adjusted upwards when the second generation is patentable in order to induce the first innovator to invest in the first place. Combining this possibility with our previous observations on patentability, *ex ante* licensing tends to allow for shorter patent lives, as the rewards can be adjusted in the licensing contract to internalise the externality before any costs are sunk. Even in the case of efficient contracting, however, patentability of the second innovation tends to require longer patent lives to ensure that innovation incentives are maintained. If a longer patent period is undesirable because of the deadweight loss, this structure of the patent can be dominated by a structure with strong novelty and non-obviousness requirements for cumulative innovations. Under this policy, relatively small steps (clear follow-ons) would often not be patentable[60].

One way to think of cumulative innovation is to think that innovations now move up a quality ladder so that improvement innovations can make the earlier innovations obsolete. A process of Schumpeterian "creative destruction" occurs as we move up this ladder. This process, however, can render the statutory patent life irrelevant since innovations are eclipsed before the statutory length of protection is reached. It is not clear, then, that statutory length can have the same "scaling" function that we have attributed to it. In fact, it is no longer clear that we can make the strict separation between length and breadth of protection as independent policy tools that we did before. Now, small leading breadth is no longer consistent with an infinite stream of monopoly rents. While the notion of lagging breadth is well-defined, leading breadth and the statutory length of protection combine to determine an expected effective length of protection. Another way of seeing this is to say that the statutory length may scale up profits if improvements are slow to emerge, but cannot necessarily be relied upon to create such scaling if improvements come quickly. One might then need to rely on other tools, such as leading breadth, to do this.

A second comment on the cumulative innovation models we have reviewed is that the previous models assumed that the "roles" of first and second innovator were clearly assigned. In point of fact, the same firm may sometimes function as a follow-on innovator and may sometimes be the first innovator. The distinction between first and second *inventors* then becomes blurred even if the distinction between first and second *innovations* is clear. Despite the prominence of bargaining, the role of patent design may not, then, be to transfer profits from one "type" of innovator to another: all firms are potentially of all types. Instead, the aim is to balance *total* profits to innovation for each innovator against deadweight loss. In this sense, we move back towards the trade-offs found in the single innovation literature.

O'Donoghue, Scotchmer and Thisse (1998) examine such a quality ladder setting where each firm can take on both leader and follower roles and where statutory length and leading breadth interact to jointly determine patent rewards. In the phrasing of Hopenhayn and Mitchell (2001), the patent system in a quality ladder framework establishes a clock that is running on monopoly rights for lower quality firms as well as a promise of rights for the firm currently holding the high quality innovation. Suppose that the magnitude of each innovation's quality improvement over the

[60] There are limits to this argument, as delays in research plus a short patent on the first innovation mean that the second innovation does not infringe for at least some of its life. Furthermore, if it is not clear that follow-ons depend heavily on the first innovation, the externality argument gets weaker. In both these cases, the argument for patentability of the second innovation gets stronger. See Scotchmer (1996) for more discussion.

previous frontier technology is v, and this also indexes the profitability of the improvement until it is supplanted by the next innovation or the patent expires, whichever comes first. Then each improvement generates a potential value to society of $v/r - c$, where c is the development cost of the improvement and the improvement generates value forever discounted at rate r. On the other hand, if creative destruction occurs each period, the innovator only earns v for one period, after which the product becomes "obsolete". This private reward may be insufficient to cover development cost c. We have insufficient innovation incentives since each innovation creates value that benefits society forever, but the innovator collects this value only for a single period. Indeed, even if statutory patent life is infinite, an innovator will never have full incentives to innovate as long as creative destruction occurs at some rate.

Define leading breadth in this framework as a quality margin, k, such that if an improvement possesses a quality margin less than this it infringes the patent. Now consider alternative patent protection designs in this framework. Suppose first that for the duration of a patent, all improvements infringe (so that leading breadth is infinite). Ideas for improvements arrive at some rate to independent agents. Since *ex ante* licensing is permitted, all improvements where the net value is positive will be made but will incur a licensing fee that splits the surplus between the improver and the holder of the infringed patent. If patents last a number of periods, T, each innovation v would then earn "direct" discounted profits $X(T)v$, but would also earn licensing revenues from improvers -- and would result in licensing *payments* to earlier infringed patents -- during its life. Call net licensing revenues $L(T,h)$, where h describes the history of previous quality improvements. Hence, the innovation constraint now becomes $c \leq X(T)v+L(T,h)$. We can define $v(T)$ as the schedule of quality steps that satisfy this constraint with equality for different patent terms: this set of steps defines the marginal innovations that will be invented.

Consider now the alternative design where all patents have infinite life but limited breadth so that only creative destruction causes them to "expire". Any improvement within margin k creates profits until it is supplanted by a non-infringing improvement: that is, an improvement lying outside margin k. Infringing innovations will be created as long as the profit surplus they create is non-negative due to *ex ante* licensing. Hence, during the period of protection, the patent holder earns revenue composed of direct returns plus licensing fees and payments. Until a non-infringing innovation is discovered, this innovator will remain the market incumbent, earning the revenue stream. Formally, if the discounted profits of a patent lasts some set of periods, t, before being replaced by creative destruction, but this duration t is distributed according to a Poisson process with arrival rate Γ (reflecting an uncertain research process), the authors assert that the expected net discounted profit from any improvement, v, is $\frac{v}{r+\Gamma(k)} - c$. Define $v(k)$ as the quality step that just sets this expected net discounted profit equal to zero for patent breadth k. All innovations at least as large as $v(k)$ will be created, even if they infringe and so require a license, so that $v(k)$ is the marginal innovation under the alternative regime.

The marginal innovations are not necessarily the same under the two policies, giving rise to differences in the rate of innovation and the research expenditure under the two protection regimes. As protection increases towards infinity on both dimensions, breadth and length, however, the rate of innovation tends to approaches the social optimum. Hence, the flavour of the result is similar to earlier cumulative innovation models where there is a tendency for very strong protection to be optimal.

As we have noted, however, the two policies we have just considered are not equivalent. To induce the same rate of innovation the first policy is associated with a shorter effective patent life. In the first policy, the binding dimension of patent protection is its statutory length so that a patent holder has claims on future innovations. The same level of investment incentive can be created with relatively short protection in this case. Rewards to the innovation are high, so statutory protection can be short because the total reward quickly surpasses the cost of innovation. On the other hand, broad short patents potentially create deadweight loss by concentrating the rights to use innovations in a few hands with little "close" quality competition, recalling the single-period models. In the second policy, the binding dimension is patent breadth, so that follow-ons tend not to infringe, and effective life must be determined as a consequence of breadth. In other words, the claims on future generations of innovations are quite limited in this case. To achieve the same initial investment incentive, the effective patent life must be adjusted to be longer for narrower breadth and shorter for larger breadth[61]. When demand is inelastic, the lower R&D costs that come with the latter policy make it preferable since the longer patent period does not create deadweight loss. If there is a deadweight loss associated with the period of patent protection, then the first policy can be better as patent protection is shorter.

Translating the policy in this paper into patent statutes is, as usual, tricky. One way to think of operationalising the policy of very large leading breadth but short statutory patent length is to think that every technology that relies on prior art (i.e., any patent that cites an earlier patent) infringes that prior art for the duration of those patents. A way to think of the policy of narrow leading breadth but long statutory length patents is perhaps by applying a strict interpretation of the doctrine of equivalents (where equivalence is in terms of quality step), which can have the effect of granting very limited scope to patent claims beyond what is actually enunciated in the patent claims themselves[62].

As O'Donoghue (1998) points out, the above result on leading breadth relies on a well-functioning licensing market. When efficient licensing is not possible, (perhaps because it is difficult to identify subsequent innovators as a practical matter or because transactions costs are high, may obtain a stronger argument for the importance of a *patentability* requirement to obtain optimal innovative behaviour in such a quality ladder framework[63]. Note that we considered infringement standards in the O'Donoghue, Scotchmer and Thisse framework, but we did not consider whether

[61] Horowitz and Lai (1996) anticipate this point using patent length and the frequency of "creative destruction" to determine the incentives to produce a "big" innovative step. Their work interprets length as statutory length, while O'Donoghue *et al* (1998) make it clear that effective length can be determined by either statutory length or leading breadth.

[62] For a recent legal decision in this area, see *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722 (2002).

[63] See Gallini (2002), Merges and Nelson (1990) and Heller and Eisenberg (1998), among others, for a discussion of impediments to licensing. Comino et al (2007) show that both early innovators and followers may benefit from decreased breadth when licensing is inefficient because the first innovator cannot observe whether follow-on innovators have already undertaken R&D activity, which decreases licensing's effectiveness as a tool.

follow on innovations should be separately patentable. O'Donoghue reasons that, if it is assumed that an unpatentable innovation earns no profits, then there is no point in targeting unpatentable innovations in research. As a result, a larger patentability requirement can induce firms to target larger innovations -- since these are the profitable ones. If these big steps take longer to accomplish, then this policy comes hand in hand with increased rewards to innovation since larger steps tend to prolong the effective period of incumbency. As a result, the reward to research can be increased by a tough patentability requirement. In other words, patents promote, but also *retard* research by effectively discouraging innovations inside a quality threshold, and so a patentability requirement can modify the chosen step size on a quality ladder. Imposing a patentability requirement so that firms target innovations larger than the social optimum can, in fact, improve dynamic efficiency. This is the case because firms tend to invest too little when they can be eclipsed by followers. The patentability requirement tends to increase R&D incentives, which has a first order effect on welfare, while the adjustment to the innovation "step" has a second order effect when that step is close to the social optimum. The point is, then, that patentability requirements have "bite" if licensing functions poorly. As an empirical matter, Hall (2007) notes that there is evidence that firms direct their research towards patentable rather than unpatentable subject matter so the underlying assumptions of the model seem to receive some support[64].

In the cumulative innovation papers considered so far, it would be best if a single firm had the ability and resources to carry out the entire stream of innovation itself. If a single agent were responsible for the entire stream of innovations, the externality would be internalised. This is the nub of the tendency for these models to favour very strong protection for seminal contributions. However, this single firm benchmark neglects the potential benefits and costs of having several potential innovators "race" for the rights to a given "idea". If "ideas" are not public knowledge, then these potential benefits are irrelevant: each innovator pursues his or her own "ideas" without the fear of being beaten to the punch by a rival. If, however, research ideas have a significant public knowledge dimension, then the potential benefits and costs from "racing" cannot be neglected in patent design. This issue is addressed by Denicolo (2000) who points out that, because of potential duplication of efforts and the incentives to pre-empt, the private market may over-provide innovation. It may be better, then, to reduce the reward to innovation. This effect can, then, dampen the optimality of heavily rewarding firms that create seminal innovations that was present in earlier models. In fact, once we introduce the possibility that firms race for innovations, aligning the private and the social reward to innovation without considering duplication may be the wrong policy as the losses from duplication may be very large.

[64] In a model similar to O'Donoghue (1998), Hunt(2004) finds that the inventive step requirement for patentability that maximises the rate of innovation is at an intermediate level. While increasing the inventive step requirement makes the marginal discovery unpatentable (so that R&D expenditure is "wasted"), it prolongs the reward to patentable steps since discoveries exceed the patentable threshold less frequently. This increases the incumbency period. If exogenous parameters are such than an industry tends to invent frequently, increasing the inventive step has a large marginal effect on rents as they are discounted little. Hence, "high tech" industries -- with frequent innovation – optimally require a more stringent inventive step than those with infrequent innovation.

Maurer and Scotchmer (2002) address the issue of racing by arguing for an independent invention defence to patent infringement (currently available for copyrighted material). In other words, they suggest that racing concerns could be addressed by another policy tool: that of allowing a firm that has conducted duplicative but non-imitative effort to commercialise its invention. This has the benefit of reducing deadweight loss by introducing more competition into the final market. It also puts a "cap" on earnings and so reduces entry into the race, which dampens duplicative expenditure. If the social benefit of reducing this deadweight loss exceeds the negative impact on the innovator's incentives to invent, then it can be socially beneficial to allow independent inventors to co-exist in the market. Of course, determining whether invention was truly independent or simply imitative could be a daunting task.

Summarising the papers examining patent length and breadth in the cumulative innovation case, one can suggest the following conclusions. First, there appears to be a relatively strong argument for protection from literal imitation (large lagging breadth). Leading breadth has more qualified support: its benefits rely on the assumptions one makes about the scope for licensing. If licensing is fully flexible and efficient, then a strong argument for leading breadth exists. If licensing possibilities are restricted, then a much more limited case for leading breadth can be made. Strong patentability requirements receive some support when licensing does not function well. When duplicative investment as a result of racing is taken into account, there is an argument to be made against very large rewards for any invention. Indeed, racing considerations generally limit the argument for strong patent rights.

Complementary Innovation

Lemley and Shapiro (2007) suggest that it is not just the cumulativeness of innovation that creates a difficulty in allocating an externality. If patents are complementary, with synergistic benefits such that the sum of the patents adds up to more than the separate parts (for example if a product is made possible only by the combination of the patents) then we can also get socially incorrect levels of innovation investment. The reason is that there is both an externality and an investment coordination problem that did not exist before. In the case of cumulative innovation, there was a (positive) externality that ran only one way, from the first innovation towards the follow-on. Now, the externality runs two ways, as each innovation is a necessary "piece of the puzzle" in the final composite good. Further, one innovation does not necessarily completely precede the other in time. That is, when innovations cumulate, the follow-on investment does not begin until the first innovation exists. In the case of complementary innovation, however, all investments could potentially occur simultaneously. It could, then, be possible that multiple innovation equilibria exist: it could be an equilibrium for all innovations to be created or for none of them to be. Hence, as a result of the two-way externality and this difference in timing, there is a pure coordination problem in investment to be solved that was not present before.

If pooling a variety of patent rights is necessary to create a final product and the licensing transaction is costly, Heller and Eisenberg (1998) make a general point that when multiple, separate, rights holders must be brought on board to create social value, innovation may be underprovided due to transaction cost considerations. They identify this as a "tragedy of the anticommons", in contrast to the more classic tragedy of the commons. While this issue existed in the cumulative innovation case, it may be more severe in the case of products that read on a wide number of patents in a variety of fields simply because the relevant patents may not be filed over time but may instead be simultaneous and so in force for a long time. If licenses are not negotiated, then there is a potential for an innovative good never to make it to market in the first place, resulting in social loss[65].

Shapiro (2001) examines formally the case where multiple rights owners contribute to a new product or process, creating a "patent thicket" that a new product could potentially infringe. Shapiro draws an analogy to the "Cournot complements" problem where a manufacturer must purchase n essential inputs from n distinct monopolists. Suppose that each $i = 1,\ldots n$ separate firms owns a patent that is essential to the production of a final product to be sold on a competitive market. Each firm sets a per unit royalty, r_i, for its patented "input" and each patented "input" is produced at marginal cost o_i. The final good price, p, will be composed of some manufacturing cost for the assembler plus the sum of all the royalty payments charged for access: $p = c + \sum_{i=1}^{n} r_i$. If each of the royalties is set independently and non-cooperatively, then for price elasticity of final demand ε the mark-up of the price over the marginal cost of "input" production will be:

$$\frac{p - (c + \sum_{i=1}^{n} o_i)}{p} = \frac{n}{\varepsilon}$$

which is n times the standard monopoly mark-up. The final price of the manufactured good is higher under this vertically separated structure than it would be if a single vertically integrated firm provided (all) the inputs and output. It is also higher than the price that would be charged by a competitively organised final product market, which purchases from a single, monopolistic supplier of all essential inputs. The profitability of the innovation as a whole falls because individual firms fail to internalise a (negative) pricing externality. As a result, there is a socially undesirable reduction in the research incentive.

Since the Cournot complements problem penalises members of the industry as well as consumers, one would expect to observe institutions to have arisen to limit this behaviour. Where high technology products rely on technological standards that are composed of multiple essential patents owned by different parties, the patents are often required to be licensed at "reasonable and non-discriminatory" (RAND) royalties. While this can be seen as a way to limit royalty overcharges, Schmidt (2008) comments that it would be very difficult to implement vague words like "reasonable" in any systematic way and, indeed, quotes Swanson and Baumol (2005) who state that "It is widely acknowledged that, in fact, there are no generally agreed tests to determine whether a particular license does or does not satisfy a RAND commitment".

---

[65] Walsh, Arora and Cohen (2004) and Walsh, Cho and Cohen (2005) suggest that at least in the case of research tools in the biomedical industry, the "anticommons" problem may not be very severe empirically.

The Cournot complements line of reasoning we have just developed clearly yields social efficiency arguments for various policies. Schmidt (2008) examines horizontal merger among patent holders as a remedy. He also allows for market power downstream so that a double-marginalisation problem exists on top of the complementarity problem. Under sufficiently flexible licensing contracts (such as two-part tariffs), merger solves both inefficiencies. He notes, however, that Layne-Farrar and Lerner (2008) find that all the patent pools they investigate used linear royalties. When he restricts contracts in this way, horizontal merger among inventors continues to perform well, but vertical integration does not. The reason is that each vertically integrated entity does not internalise the externality it exerts via its royalty rate charged to other (vertically integrated) entities. To the contrary, by raising the royalty, each entity can raise rivals' costs. Further, each entity suffers from some double-marginalisation for the patents it must buy in.

If one assumes that the set of patents in the "thicket" is not "fixed", but is accumulated over time due to continuing research, Noel and Schankerman (2006) hypothesise that a reasonable reaction to the Cournot complements problem could be for firms to accumulate large patent portfolios. Indeed, they find some empirical evidence for excessive incentives to patent in order to "hoard" in the software industry[66]. Related work by Arora et al. (2001), Hall and Ziedonis (2001), and von Graevenitz et al (2008) finds that the recent growth in patent applications can be attributed to defensive use[67] of patents in "complex" industries – those where patent thickets are present.

Alternatively, allowing complementary patents to be traded as a "package" for a single price rather than traded separately could yield gains. Hence, we might wish to treat patent pool agreements – agreements among multiple patent holders to aggregate a set of patents among pool members or license as a package to non-members – leniently when they involve complementary patent rights[68]. Cross licenses could serve the same purpose[69].

Not all industries are equally susceptible to complementarity problems. Cohen et al. (2000) classifies industries according to whether they are "complex" – so that value is derived from complementary components -- or "discrete". If this is the case, targeted industrial policy towards patent pools or merger could address the complementarity problem. Alternatively, one could think of the complexity of an industry as the result of patent design: if patents are granted very narrowly, then many complementary "bits" would necessarily contribute to almost any product. Perhaps the appropriate

---

[66] They hypothesise that a larger patent "arsenal" also strengthens the bargaining position of an inventor and reduces transaction costs as the number of potential negotiations fall. Dewatripont and Legros (2008), in an analysis of patents' contributions to standards, appeal to a version of a Shapley value to justify the relation of bargaining strength to the proportion of patents owned.

[67] This defensive use can include litigation concerns, which will be discussed below.

[68] See Schmidt (2008) for a summary and comments on recent US policy moves towards a "rule of reason" approach to whether patent pools must contain only complementary patents or whether substitutes can be included as well. Layne-Farrar and Lerner (2008) give a history of patent pool policy in the US.

[69] For examples of patent pools and their diversity (from mega pools comprising a broad-based governance structure for huge numbers of patents to small pools that amount to no more than a few multilateral contracts establishing a way to consolidate patent rights and a rule to divide up licensing revenues) see Merges (1999).

policy response is to make patents broader, then, to reduce the cases where complementarity issues arise.

Both of these solutions could be hasty, however. The reason is that both assume that the degree of complementarity is not a choice variable for the producers. Lerner and Tirole (2004) take the opposite tack suggesting that, while the Cournot analogy provides a good starting point, it is often difficult in practice to pinpoint whether a patent is a complement or a substitute for another. The complementarity of patents may be less an unchanging "objective" characteristic of the patents than a characteristic of how a particular manufacturer optimally decides to combine technologies into a final product. Worse, these characteristics may change over time as technology and its applications progress. This could call our policy responses into question.

To sketch Lerner and Tirole's argument, consider a case where users can purchase patent "inputs", supplied by n upstream owners, each of whom owns one patent. Users combine these patents in various ways to create a valuable product. Users will do this in a surplus maximising way, which necessarily takes into account the input cost (i.e., the price at which the patent is licensed). Users may create surplus either by using a subset, m, of the n available patents or the entire available set. More precisely, a user can combine patents to create value $\theta+V(m)$, where V is an increasing function of the number of patents actually used, $m \leq n$, and the values θ are distributed according to some cumulative distribution function, G, over the user population. Hence, just using a single patent creates value, but the set of all patents creates even more value and further, users are heterogeneous in how much value they derive from the final "product".

A patent owner can license these innovations to users at a price, P. User demand will be determined by the value extracted from the patents, $\theta+V(m)$, net of the price at which they are sold, P. For low P, it may very well be the case that the net value extracted from the patents is highest when all n patents are bought. This would be the case when the price for each patent, p_i, is less than the value added of the m^{th} patent, $V(m) - V(m-1)$, for all m. Note that, when all patents will be used, a rise in the price of one patent will tend to decrease the attractiveness of the final good as a whole because its price will rise. Hence, a price rise for patent i decreases the demand for other patents. This means that the patents are demand complements at low prices: the rise in the price of one causes a fall in the demand for another. On the other hand, it is possible for the prices of a patent to exceed its marginal contribution over some base number of patents. In this case, only a subset of patents will be purchased and combined to create end value. For example, if $n = 2$ and each patent is priced above the contribution of a single patent, $\theta+V(1)$, but below the marginal contribution of the second, $V(2) - V(1)$, then only one of the patents will be purchased and used. In this case, a rise in the price for one patent causes the demand for the other patent to rise because use switches to the cheaper of the two. Hence, the patents are substitutes.

In setting a license fee, then, a patent owner needs to take into account two effects. First, she needs to think about whether the patent will be retained in the "basket" of patents that are purchased. Second, she needs to take into account the effect of her own fee on the final price of the good that the patents are used to create (and hence the final demand for the "basket" of patents). If the second effect is dominant, then under non-coordinated pricing, each patent holder exerts a positive externality on other patent holders when she lowers her price since she raises demand for the entire "basket" of patents. Price for the basket will fall when this externality is internalised, so that coordinated "pool" pricing reduces user price and raises welfare in general. This argument recalls the Cournot reasoning, above. On the other hand, if the dominant effect is the first, noncoordinated pricing may induce an incentive to lower each patent price so as to "steal business" from other patent holders. This can create welfare gains to noncoordinated pricing over pool pricing if patents are sufficiently substitutable. Hence, the recommendation for patent pools is more nuanced: we only want to be lenient when an endogenous "complementarity effect" (and not some exogenously determined "complementarity characteristic") is dominant.

Disclosure issues

As we said earlier in the chapter, a major function of the patent system is to disseminate information. One could think of this as transforming private ideas and their embodiment into public knowledge by means of the patent disclosure requirement. Hence, the degree to which ideas are "private" is a policy instrument of the patent system as well as a choice variable for firms that can select between patenting their innovations and exploiting them as trade secrets. The degree to which information will actually be revealed in a patent system and the degree to which it will be withheld in a secrecy system is, however, debatable as we will see below.

Maurer and Scotchmer (2006) emphasise the coordination role of the disclosures, arguing that one of the benefits of disclosure comes from informing the inventing community generally of who is working on what, which results have been obtained and which have not in the same way that publishing serves the academic community. In a sense, the disclosure requirement of patent law creates a "public repository of knowledge"[70]. If a first inventor could easily identify the best qualified "next" inventor, she could disclose any relevant information to subsequent inventors privately for a fee, thereby profiting from the increased efficiency of the research path. Without this information -- with unanticipated applications of technologies coming from unlikely sources -- a public repository of knowledge may be the most efficient way to allow those with the skill and creativity to make the next step to actually contribute. To the extent that licensing actually occurs, allowing technologies to work together, the "public repository" should also allow inventors to specialise in their area of technical competence. This coordination benefit becomes more significant the more efficient the licensing market is.

Denicolo and Franzoni (2004) add to the argument for the coordination benefits of the patent library by suggesting that the patent disclosure can reduce – perhaps unintentional -- duplicative research effort. Whether the disclosure is a tool that actually publicises this information in a form that can be interpreted and accessed readily is, perhaps, more debatable. Bessen and Meurer (2008) have argued that, in fact, many infringement cases are inadvertent. This could suggest that potential the coordination benefits of a library are not being realised fully.

[70] Aoki and Speigel (1998) suggest that the recent move in the US towards earlier disclosure may have significantly sped up the development of technology by improving the available research base.

Even if competitors are already aware that a potentially profitable investment opportunity exists, the information in the disclosure can affect the nature of the race towards discovery within this general area. Disclosures make the information structure in an R&D race a choice variable for the participants. For example, suppose that Ms. A possesses an innovation that is secret and that gives her a hint about how a future innovation could be designed. Ms. A knows that Mr. B. is working on the same problem, but has not yet obtained such an intermediate result. If Ms. A discloses this information via obtaining a patent, she gives up her lead in the R&D race. This can create a large disadvantage to the disclosure system for users. On the other hand, Ms. A. can commercialise her innovation without fear of imitation due to the protection the patent affords on her intermediate step. In a system where taking a patent is voluntary and secrecy is always an alternative means of protecting the gains to innovative effort, the trade-off faced by Ms. A. suggests that not all innovations will be patented and disclosed. Only those innovations will be patented for which the trade-off goes in the direction of large gains to commercialising under patent protection and little loss in terms of an R&D race. Hence, a patent system with disclosure only ensures that some innovations may be disclosed, not that all innovations are disclosed. This, too, can hamper the coordination role of the patent system.

If disclosure has a benefit, perhaps society would be better off in a system where secrecy is not an option. Aside from the difficulty of enforcing such a policy, there are reasons why allowing firms the option of patenting could be beneficial. If firms have an observable choice between secrecy and patenting (so that it is possible to observe that a firm is keeping a secret, but it is not possible to know what the precise nature of the secret is), then the act of patenting can have signalling value. An early contribution by Horstmann et al (1985) takes the view that the simple act of patenting signals information accrued by the inventor during an R&D stage. If the information thus revealed makes imitation around the patent more profitable for a competitor, the propensity to patent falls. Forcing full revelation (so that the signalling function is compromised) is not necessarily welfare improving due to welfare losses from increased imitative R&D expenditure[71].

Further, we have already seen that Anton and Yao (1994) show that under certain conditions a limited amount of revelation will occur under secrecy, as inventors reveal their innovations privately to a limited set of licensees. Hence, a system of secrecy will be associated with some disclosure. Anton and Yao (2004) examine which *types* of innovation may tend to be patented rather than kept secret in a signalling framework. Their model shows that it is the smaller innovations that will tend to be patented (and disclosed), rather than the larger ones. This could potentially lower the benefit of the patent disclosure since only small steps will appear in the "repository of knowledge" that the patents create. More precisely, the authors assume that the enabling information in the patent need not, in fact, allow rivals to completely duplicate an innovation. Innovators may choose to disclose a lot of information in the patent document, thereby convincing rivals that innovations are quite significant. Such disclosure triggers imitative behaviour, of course, and may result in damage payments from the imitation. Innovators could also opt for trade secrets that disclose

[71] More recently, Langinier (2005) develops this line modelling.

very little but also give no rights to damages in the case of imitation. A separating equilibrium exists where small innovations are patented, fully disclosed in the patent document, and are not imitated; large innovations are kept secret and are not imitated because no information was disclosed. This model assumes a weak enablement requirement so that partial disclosure is possible. The informational requirements apart from the specific information in the enabling requirement are quite large: it must also be possible to know what *proportion* of a total amount of information was disclosed in order to derive the equilibrium in the first place. The basic point they are making, however, is that as long as secrecy comes along with sufficient control, it can generate selective disclosure. Further, as long as patenting is a choice and not an obligation, only certain types of innovation will be disclosed via patents. If the patented innovations are not the most socially valuable types to disclose, then the patent disclosure does not function optimally[72].

The enablement requirement, while clearly linked to how much information is disclosed in the patent system, is not the only tool that affects the disclosure function. How should other aspects of the patent be designed to obtain the most out of disclosure? Scotchmer and Green (1990) examine the novelty requirement for patentability and how this can be managed to promote disclosure. Disclosure may be well-served by a weak novelty requirement for patentability where even small improvements can be patented. If the uptake of patents on these intermediate steps is good, scientific progress building on known art can be rapid. This advantage is undermined if firms do not choose to patent the interim innovations in order to avoid giving away valuable information; however, a strong novelty requirement does nothing to help resolve this problem, as no further disclosure will occur under this regime. The novelty requirement also affects the incentives in the patent race, however. A weak novelty requirement could have the effect of ensuring that the market is populated with products that are relatively close substitutes and, hence, are not very profitable. While a strong novelty requirement could lead to slower discovery by any one researcher, the larger reward that it promises to those who remain in the race could lead to increased entry into innovative activity. Entry could ultimately speed up final discovery. Patenting is voluntary, however, and the fact that the weak novelty requirement opens the *possibility* that close substitutes would be provided does not *ensure* that patenting occurs. Scotchmer and Green show, to the contrary, that firms choose to suppress the interim discovery precisely when profits would be eroded. Hence, the weak novelty requirement does not necessarily lead to low rewards. One can, however, make a signalling argument for a strong novelty requirement: if the novelty requirement is weak, a firm can infer something from the very suppression of an invention when patenting was a viable option. The inference that an invention has been discovered but suppressed can discourage innovation investment by a rival who thinks she has fallen behind in the race[73]. If the novelty requirement is strong, there is no option to patent so there is no signalling value in the

[72] Anton and Yao (2002) examine the case where ideas can be partially disclosed so that some information can be revealed, but other information can be traded privately as unpatented "know how". In their 1994 paper, where partial disclosure was not possible, the licensing contract offered by the purchaser to the inventor only had to eliminate the incentive to disclosure the information to a third party. In Anton and Yao (2002), disclosure can also be used to signal the extent of other – undisclosed – knowledge that the seller possesses. The undisclosed knowledge will be bid for by competing buyers and not all knowledge may be disclosed in the equilibrium of interest.

[73] See Gill(2008) for more discussion of strategic disclosure as a tool to make a competitor drop out of an R&D race.

observation of no patenting. Of course, a weaker enablement requirement -- so that information in the patent disclosure need not be very complete -- could undermine this argument as the overall disclosure benefit as well as the signalling value of patenting would fall[74].

In related work with a somewhat different legal interpretation, Baker and Mezzetti (2005) and Bar (2006) develop the idea that the disclosure of intermediate steps in a patent race affects prior art. This means that disclosure of intermediate steps can affect the patentability of subsequent innovations because the subsequent innovations must be novel when held up against this prior art. Disclosures may be optimal in their framework even if they are not accompanied by patent protection. As prior art is built up of any public information (patented or not), the leading firm must make a greater improvement to obtain an innovation viewed as sufficiently novel to patent. Hence, laggard firms may wish to disclose in order to prolong the race towards a prize that gets ever farther away: the disclosure buys them needed time to attempt to pull ahead in a stochastic R&D race framework. The decision to patent, then, comes hand in hand with a decision to make the R&D stage a more complete information race. The exact interaction between the information, the exclusive rights, and the patentability criteria determines whether firms race more or less intensely and so whether discovery comes sooner or later.

Matutes et al. (1996) focus on the distinction between a patent's disclosure of both the embodiment of an innovation and the idea underlying it but the patent's protection applying only to the embodiment. Further, in the other models we have reviewed, it is supposed that all parties realise that a secret is being kept and all have a rough idea of its nature so that there was a lot that could be inferred from "no information". Matutes et al. take the opposite tack of assuming that until the seminal information is revealed, competitors have no clue of its existence. Hence, "no information" does not act as a signal. Instead, they focus on when to disclose when disclosure both allows the innovation to be commercialised under patent and also initiates a race for the remaining unpatented applications of the underlying idea. Sketching their "waiting game", let it take one unit of time to develop each profitable application (claim) of the basic insight. Then an innovator has an incentive to keep the insight secret by waiting a period of time before introducing any of its applications, as this postpones the time others realise that a fertile insight is available to be built upon, and so start developing applications of their own. In other words, once the "cat is let out of the bag", m potential entrants will potentially start to develop any applications that have not already been developed and protected by patent by the first innovator. Hence, there is a positive externality of the disclosure that is not internalised by the first innovator: the first innovator will use trade secrecy to postpone this race to grab applications, even though waiting is socially harmful because it delays commercialisation of the applications. From the initial inventor's point of view, the impatience to commercialise the applications that already have been developed creates an incentive to patent early to weigh against the incentive to prolong the period of development "in secret". More precisely, if A potential applications exist and the initial innovator

[74] In a related point, Aoki and Spiegel (1998) have suggested that the recent move in the US towards earlier disclosure can significantly speed up the development of technology by improving the available research base. Shapiro (2004) suggests that early disclosure may have been associated empirically with less strategic activity that could be undesirable, such as so-called "submarine" patents.

waits λ units of time before disclosing, then when each application earns discounted profits of π the innovator would choose to wait according to:

$$\max_{\lambda} \ \pi(\lambda e^{-\lambda r} + \int_{\lambda}^{\frac{A+(m-1)\lambda}{m}} e^{-rt} dt)$$

They investigate the extent to which limiting this kind of waiting to access the patent system in the first place can be addressed legally by granting claims on applications that are not yet fully worked out. This has been termed granting a "license to hunt" by means of a patent. Granting a limited license to hunt increases the incentive to disclose the first step because if creates an entry barrier – a zone of exclusivity -- into a set of potentially profitable innovations[75]. Indeed, under this lens, leniency on large patents, with many (non-overlapping) claims, submitted at an early stage has a welfare benefit of speeding up useful disclosure[76]. Of course, the disclosure is assumed to be useful here: potential researchers face no difficulty in wading through and digesting reams of claims[77]. The paper suggests that delay that facilitates "hoarding" may occur in the case where accessories could follow on to an original "platform"[78].

Hence, both the enabling disclosure in the patent and the act of patenting *per se* carry information that affects imitative and innovative behaviour. While the positive externality conferred by the enabling information has social benefit, private parties may be expected to disclose less than the socially optimal amount. This limits the value of patents as a "repository of knowledge". Secrecy does promote some limited disclosure, although the disclosure may be private to only licensing partners. On the other hand, the standards of enablement, the novelty requirement, the role of previous patents in the definition of prior art, and the leniency in granting claims on prospective applications are all available tools to affect the amount of disclosure actually obtained by the patent system. More complex strategic reasons to decide to patent or not also exist and can be used to manipulate the behaviour of rivals. This latter function derives from the optional nature of patenting, and the existence – in some cases -- of commercially viable alternatives.

E. Alternatives to a Patent System: Optimal Procurement of Innovation

---

75 The claims in new and quickly developing fields can be rather speculative, in particular in some pharmaceutical areas according to Bigdoli (2009), ch. 216 "Innovation and Intellectual Property".

76 The idea of giving prospective protection on claims that are not fully developed recalls the work of La Manna (1994), discussed above, where there is a benefit to reserving territory to a single patent-holder before investment occurs, although diffusion benefits are not a concern in that model. Kitch(1977) identifies a "prospect function" for patents when they are granted early in development. See Merges and Nelson (1990) for a discussion of this view.

77 Chen and Iyigun (2006) incorporate concerns about delay in patenting and disclosure into a model of economic growth. Duplicative research expenditure is less of a concern in their framework, so imitation will be observed in their optimal patent design, in contrast with most of the papers reviewed here.

78 Noel and Schankerman (2006) finds evidence for hoarding in complex industries. No complementarity is required to obtain hoarding in the Matutes et al model, as the accessories are assumed to be independent "pots of gold" for whoever innovates them.

The discussion of the preceding section took for granted that some system rather like the current patent system would be used to generate innovation incentives. We considered modest modifications of this system, using existing tools such as statutory patent length, infringement standards, a patentability requirement, the enablement requirement, and the interpretation of claims to achieve that goal in the most efficient way. If we were to start from a blank slate, it is not clear, however, that patents as we know them would be our chosen optimal scheme to "promote the progress of science and the useful arts". Indeed, many other schemes have been and are still used to achieve this goal[79].

Wright (1983) develops the point that a patent-like system might not be the best mechanism in a fully optimised model of innovation incentives. Instead of using the monopoly mechanism to create a reward for innovation, the state could instead simply award compensation in the same amount directly to the innovator and obtain a welfare gain. In other words, whatever the reward available through the patent system, transferring that reward to the innovator as a lump sum rather than as a result of a market distortion achieves the same innovation incentives with less deadweight loss. He refers to this sort of payment made by an authority and conditional on the delivery of a completed advance as a "prize". Hence, if the value and cost of the innovation are publicly observable, and if the funding of the prize is relatively non-distortionary, then the prize system will dominate the patent as an incentive mechanism. Wright goes on to suggest that subsidies could dominate patents as well. Competitive bidding can be used to contract out the research before its completion to ensure that only the most efficient researchers are used. While this removes the normal racing incentive, this system can achieve higher welfare than a patent if timely innovation can be induced via performance requirements.

Information is rarely this good. In particular, it may be more realistic to assume that the authority awarding the prize has less information about the value of any candidate innovation, let alone its cost, than the innovators themselves. In the presence of asymmetric information, Wright shows that any one of the three mechanisms -- patents, prizes or contract research -- could be the best mechanism. Patents have the advantage that they delegate the decision of which investments to put forward to the "informed party", the inventor. If the inventor is the one who knows which investment will generate the most value, but the sponsor does not (without incurring a cost), this represents an improvement over a prize system that would require the government to "pick winners"[80]. This gain can outweigh the deadweight loss associated with the patent system. Further, a prize has a drawback of its own in the case where innovative activity can be conducted by many parties since the prize does not limit entry. Hence, it suffers from generating excessive research expenditure due to the "common pool problem" that was discussed above as part of the reward theory. One alternative to reduce this incentive for duplicative expenditure is to reduce the size of the prize. If the funding body has poor information on the value or cost of the

---

[79] See Scotchmer (2004) for an in-depth review and history of other schemes to promote innovation, including prizes, subsidies and direct procurement.

[80] Shavell and Ypersele (2001) show that a combined reward cum patent system always dominates a pure patent, taking as an example a case where innovations can be ranked by their value and the prize is set so as to induce the lowest innovation "type" (value) to switch from patent protection to the prize. This reduces deadweight loss and does not suffer from a problem of picking winners, although it does require enough knowledge about possible innovation types to be able to set this minimum reward.

research, however, this system might only elicit low value ideas while leaving the level of compensation too low to elicit research in high value output. Contracting out research potentially avoids excessive research expenditure, but it may not result in sufficient incentives to create value in the first place, precisely because it eliminates research competition. In other words, while the incentive to "race" for the patent may be too large, the incentive to invent at all for a single, designated contract researcher may be too small. The contract would need to be designed to ensure that the best idea was selected for funding and that invention incentives were maintained. The possibility of designing such a contract depends on what is observable to the funding body and on the credibility of the promise to pay for deliverables[81].

When information is asymmetric, Gallini and Scotchmer (2002) point out that designing an incentive mechanism for innovation can be broken down into three "steps". First, there is the decision problem of whether a project should be undertaken. Second, there is the delegation problem of which firms should undertake the investment and at what rate. Finally, there is the funding problem of how to reward the investments. A set of papers applying mechanism design machinery to innovation incentives have begun to address these three points.

One approach has been to incorporate a modified "prize" into the patenting system by means of patent buyouts. Kremer (1998) suggests a system that effectively awards prizes but does not rely on the planner's identifying the value of potential innovations beforehand. Let a period of time elapse during which firms hold their patents in the normal way. After this elapsed time other firms, besides the one that patents, are likely to have an idea of the private value of the innovation. This information can then be marshalled by the planner to create a reward for the innovation using a second price sealed bid auction to elicit payments for the right to the innovation from the private parties. In order to maintain the incentive to bid, with small probability the patent will be transferred to the highest bidder in exchange for the winning bid. Otherwise, the innovation is placed in the public domain. In either case, the price determined by the auction would be paid as a prize to the original patent holder by the government out of general tax funds (or some other budget associated with the office responsible for the buyout mechanism). In fact, to reflect the difference between the private and social value of the patent, the government could apply a positive mark up to the winning bid when making its payment. Whether to put a patent up for auction (or "buyout") would be at the discretion of the patent holder. If the bids are relatively low, the patent holder can refuse to sell. Clearly, this proposal is meant to complement the patent system[82], not replace it as the initial award of a patent is essential to the mechanism. The proposed approach will dominate a pure patent system as long as the administrative costs of the buyout and the cost of public funds are not too high.

---

[81] See Scotchmer (2004) for a discussion of contests for sponsored research. The more knowledge the planner has about what the target of research should be, its value, or its cost the wider the set of alternative instruments to patents. Maurer and Scotchmer (2004) discuss a variety of procurement mechanisms, including auctions, prototype contests , grants, and matching funds. Trajtenberg (2002) provides extensive discussion of how these alternative government supports have been used to advantage in Israel.

[82] Brunt et al (2008) observe that prize awards, even non-pecuniary ones, can have a large inducement effect for innovation when used in conjunction with a patent system; however, the prizes studied are not buy-outs. Rather they are awards to scientists to undertake research that can lead to patents.

Hopenhayn, Llobet and Mitchell (2006) investigate buyouts in a quality ladder framework. Formally, the authors first consider the case when an innovator's "type" – their innovative ability -- is observable. They show that the optimal patent system can take an "exclusive" form where a quality leader above some threshold ability receives patent rights while all other previous innovators' rights terminate. More precisely, suppose that a social planner can make promises of durations, *k*, in the form of a set of time periods during which no other innovation may be implemented. In each period a new innovator arises. In such a case, allocating monopoly power to the current innovator potentially curtails the planner's ability to allocate monopoly power to future innovators. Furthermore, allocating a period of exclusionary rights today may postpone the benefits one could earn from a superior innovator in the future. The planner's problem needs to take into account several constraints. First, innovators will choose the quality increment v that they target as a function both of their basic ability, θ, (their "type") and the duration of their monopoly power, k, so that we have $v(k, \theta)$ when they are assumed only to earn profits during the k periods that they are allocated. Second, the planner must keep her promises in the sense that the cumulative duration already allocated, K, must equal the duration promised to all innovators who have already implemented their innovations. Hence, if $k_p(\theta)$ is allocated to previous innovators in each period, we have $K = \int k_p(\theta) g(\theta) d\theta$, where the distribution of types is described by density function $g(\theta)$. Third, the total duration allocated in each period to previous and current innovators, $k_p(\theta)+k_c(\theta)$ cannot exceed the total discounted time horizon $\frac{1}{1-\beta}$ when discount factor β is used. Finally, the rule by which K changes in each period is that, during a single period $k_p(\theta)+k_c(\theta)$ is allocated, but at the same time a single period elapses so that $\tilde{K}(\theta) = \frac{1}{\beta}[k_p(\theta) + k_c(\theta) - 1]$ is the balance of duration that remains next period. Under all these constraints, the social planner grants $k_p(\theta)$ and $k_c(\theta)$ so as to maximise the expected present value of all future innovations, W, given that K units of time have already been allocated. This value is composed of the contribution of the innovator to social welfare, which is the contribution of the quality improvement over its development cost, chosen optimally as a result of the policy set by the planner, times the entire future duration over which the innovation will not be excluded, $\frac{1}{1-\beta} - k_p(\theta)$. Summarising, then, we have:

$$W(K) = \max_{k_p(\theta), k_c(\theta)} \int [\frac{1}{1-\beta} - k_p(\theta)] v(k_c(\theta), \theta) - c(v(k_c(\theta), \theta) + \beta W(\tilde{K}(\theta)] g(\theta) d\theta$$

subject to the constraints discussed above. Given the appropriate sorting assumptions, they establish sufficient conditions under which the optimal patent system is of a form whereby innovators who report type above some threshold $\underline{\theta}$ obtain exclusionary rights, which begin immediately upon grant without delay. In other words, when $k_c(\theta)$ is positive, $k_p(\theta)$ is set to zero. Conditional on not being replaced, a current rights holder retains unchanging protection.

Notice that the proposed mechanism addresses the problem of whether investment should occur at all as well as the delegation problem identified by Gallini and Scotchmer (2002) in the sense that only firms with types above a threshold ability find it profitable to undertake investment. It also designs a system of rewards consistent with this. The mechanism is similar to O'Donoghue, Scotchmer and Thisse's (1998) result of limiting (and constant) breadth and infinite statutory length. The framework is much less tied to legal institutions of the existing patent system, however, so it is difficult to translate the assumptions of the model into precise legal principles that are currently observed.

We have not discussed yet how buyouts enter the Hopenhayn et al model. When it is not possible to observe type θ, the "exclusive patent system" we have described can be decentralised into a *mandatory* buyout system. The buyout takes the form of a payment to the current market leader to displace her, as well as a specified buyout amount that the new leader would accept to be displaced by another. The buyout also involves a transfer fee, paid to the granting authority, and potentially varying by the innovator's type. In order to find the buyouts that result in the same duration, $k_c(\theta)$ as in the optimal patent system described above, the authors derive a revelation mechanism such that an innovator of type θ will report his true type. The payment that implements the optimal policy can be shown to be separable into two parts. The first, $\sigma(\theta)$, is a function only of the innovator's type and the second, $\gamma(K)$, is a function only of the cumulative protection, K. The authors propose that a new innovator must pay buyout $\gamma(K)$ to the existing (exclusive) patent holder and a fee $f(\theta)$ to the planner that entitles him to a buyout in the amount $\gamma(k_c(\theta))$ in the future. Since the fee entitles the patent holder to a buyout, the fee includes a component dependent on the innovator's type ($\sigma(\theta)$) and the future buyout fee. In this way, the innovator buys out the current innovator, while the fee he pays to the planner incorporates a payment that will eventually be "refunded" by a future innovator. This buyout scheme is simple in the sense that it involves only a list of fees and buyouts. The authors point out, however, that for such a system to be derived, the planner must know a great deal about the structure of the innovation system, including the cost of development and the distribution of the types. Furthermore, the model analyses only a single and definable "ladder". In point of fact, it may not be at all clear which ladder(s) a particular innovation is on. In terms of whether the patent system could be amended to look more like the proposed mechanism, the authors note that optimal pre-specified and efficient licensing payments can potentially serve some of the same functions as a buyout organised by sponsors and could effectively implement the optimal system. This would put us back much closer to the earlier literature on optimal patent design with efficient licensing.

Kremer (1998), in his discussion of his own proposals, documents many practical problems with buyout schemes. A few examples will suffice to give an idea of the difficulties of moving towards such a system. First, as emphasised by Wright(1983), the "true private valuations" revealed by this system should include business stealing effects. As we discussed above, the private value may exceed the social value when business stealing effects are present. Kremer's "mark-up" reflecting the gap between social and private values could sometimes be negative if business stealing were present. More generally, this mark-up would depend on the innovative industry's structure as well as the nature of the patented information. Referring to his own "voluntary buy out system", Kremer suggests that there is a lemons problem in the sense that firms that know that a new innovation in their own research pipeline will eclipse an existing innovation might tend to be the ones putting their innovations up for sale in order to exploit this. Hence, there is an underlying signalling problem that could affect the performance of the system and would affect its design. On the other hand, a mandatory system, such as Hopenhayn et al's formulation, avoids this by dint of being mandatory. Third, the optimal system depends on how patents interact: whether patents are complements, substitutes, or on the same -- independent -- quality ladders. We have already argued that the substitutability or general interaction between patents is not clear cut in many cases, following Lerner and Tirole's analysis. Fourth, auction-based buyout mechanisms can only be as good as the auction mechanism on which they rely. No auction mechanism is perfect. For example, the second price sealed bid auction is vulnerable to collusion amongst participants. Finally, there are knotty political economy issues associated with making a buyout system widely available to all patents, even if one takes for granted that the economic issues can be solved. For example, if it were widely publicised that frivolous patents – of which there appear to be many[83] -- were receiving payouts from the government, those paying into the general tax funds might not react well[84].

Scotchmer (1999) and Cornelli and Schankerman (1999) suggest pairing the patent system with other incentive tools, focussing on subsidies. Both start from the idea that one wishes to preserve the desirable self-selection characteristics of the patent system, while not sacrificing the advantage that subsidies have of not creating deadweight loss. Weighed against this advantage is the disadvantage that subsidies potentially encourage those applicants whose inventions have little social value to come in search of a handout.

Using Cornelli and Schankerman's presentation of the mechanism design problem, suppose that firms may be of a variety of types, where the government wishes to shift the distribution of R&D effort towards the types that are highly productive in order to minimise the social cost of producing innovations. The optimal patent policy, then, is a time of protection, T, which is a function of the announced type of the innovator, θ. In this case, we can think of θ as indexing the skill of the researcher in terms of producing an innovation (as in the Hopenhayn et al paper, discussed above) or the value of the innovation for both society and the researcher. For example, we could think of the profit from innovation as $\pi = \theta e$, where e is effort. Hence, given private information θ the firm chooses the "size" of innovation, π by setting e. The government only knows the distribution of the θ. The cost of effort is some non-concave function.

Suppose that the researcher announces $\hat{\theta}$ and the planner determines a patent length, T, and fee, f, (to be paid into the system by a patent applicant) according to the announced value, $\{T(\hat{\theta}), f(\hat{\theta})\}$. The researcher responds by choosing effort, e^*, as a function of this schedule. The welfare maximisation problem, where w denotes the

[83] Jaffe and Lerner (2006) discuss the quality of patents.
[84] If the buyout system is not self-financing, we need to take account of the effect of financing on the general economy. An advantage of patents is that they "tax" those who participate in the market for the innovation. While general tax payers under a prize/buyout system arguably gain from the elimination of deadweight loss in a system where transfers occur, the transfers would need to be established. Further, the deadweight loss benefits are more hidden than the explicit payments made to innovators.

flow welfare gain from the patent (profits plus consumer surplus only) and *d* denotes the flow deadweight loss due to patent protection (and available as a gain once the patent expires at time *T*) becomes:

$$\max_{T,f} \int_0^{\bar{\theta}} [\frac{w(\pi(\theta,e^*))}{r} + \frac{d(\pi(\theta,e^*))}{r} e^{-rT(\theta)} - c(e)] dG(\theta)$$

subject to:

$U(\theta,\theta) \geq 0$ (individual rationality)

$\theta = \arg\max_{\theta} U(\theta,\hat{\theta})$ (incentive compatibility)

Where $U(\theta,\hat{\theta}) = \int_0^{T(\hat{\theta})} [\pi(\theta,e^*) - c(e^*) - f(\hat{\theta})] dG(\theta)$

The solution of this could generally include negative fees for low type θ firms: in other words payments – subsidies -- from the government to particular types of researchers. The authors point out that this would require both monitoring schemes to ensure that low types actually innovate and public funds to provide the subsidy. Not surprisingly, the length of protection increases with type θ, in order to satisfy incentive compatibility. Hence, the length of protection, $T^*(\theta)$, (strictly) increases in θ so that heterogeneous types of researchers would generally be associated with heterogeneous protection regimes. Indeed, this optimal direct mechanism can be implemented by using either an upfront menu of patent lengths and fees or a renewal fee scheme[85].

In a more general framework, Scotchmer (1999) shows that a system that does look like a patent, a mechanism like the Cornelli and Schankerman system, is what an incentive compatible mechanism *must* look like when the economy has a single firm innovating once and where the cost and value of the innovation are not observed by the social planner (but are known to the innovator). In this system the payoff to reporting a value and cost pair such that the patent authority would ask the firm to conduct the research in the first place must at least equal the payoff to saying that the research is not socially worthwhile (individual rationality). Further, low value innovations must get a subsidy but little patent protection while high value innovations must pay a fee and get high patent protection in a way that achieves incentive compatibility. Low value innovations do not mimic high value ones as they would have to *pay* a fee instead of *receiving* a subsidy; conversely, the inventor of a low value innovation would get little value from the stronger patent protection. The subsidy instrument is set to optimally tax firms once incentive compatibility is achieved.

[85] Since $T^*(\theta)$ is strictly increasing, it can be inverted to obtain the type associated with each patent length. This can be substituted into the optimal fee schedule, $f^*(\theta)$, that is derived from the constraints on the problem, above, at the optimal schedule of patent lengths. Hence, we obtain $F(T) \equiv f^*(\theta(T))$ as the fees associated with each patent length. Alternatively, an annual renewal fee such that the sum of the renewals during the lifetime, $T^*(\theta)$, adds up to $f^*(\theta)$ would implement this solution.

The interesting result of these two papers is that the optimal system does not look "far off" from the patent cum renewal fee system that is actually observed. Specifically, the best system in these papers is a renewal system that specifies a menu of fee payments in exchange for extensions of the patent life. The patent holder purchases these extensions to patent life more readily for high value innovations, so only these patents are long-lived[86]. Further, the incentive to develop high value innovations is stronger since precisely these innovations receive longer protection. Cornelli and Schankerman show, however, that there are significant differences between the shape of the renewal fee schedule predicted by their theoretical framework and that observed in practice. First, the scheme derived from the optimal mechanism suggests subsidies for small innovations, which are not observed in practice (at the patent office, at least). Second, fees should be rising sharply over time in the optimal structure, whereas in their sample of European countries, fees actually fall. They comment that these differences should not be surprising in light of the fact that patent renewal fees tend to be set so as to finance patent offices rather than with any sort of optimal mechanism for eliciting innovation in mind. Still, these papers suggest that the "tools" exist and could be adjusted to implement a socially desirable system.

As was mentioned above, a further characteristic of these models is that the effective patent protection of different types of innovations differs: the uniform patent protection that is generally observed in practice is not optimal. Cornelli and Schankerman conduct simulations to illustrate the welfare loss of moving to a uniform level of protection from a heterogeneous system, finding that the optimal mechanism generally raises welfare 2-7% above uniform protection.

Indeed, Hopenhayn and Mitchell (2001) explore heterogeneous protection more fully in an optimal mechanism that allows patent authorities to choose the length and (leading) breadth of protection as well as the renewal fee schedule. In other words, rather than consider just fees and the length of protection as instruments, Hopenhayn and Mitchell (2001) re-introduce breadth (which can be thought of most straightforwardly in terms of quality increments in a quality ladder, but could also be an exclusion zone in product variety, similar to Klemperer (1990), in their general framework) as an instrument to determine how all three of these tools can be combined optimally by social planners. Their paper brings the patent design literature "full circle" in the sense of bringing the tools explored in the earlier literature into a mechanism design framework. The paper also derives the result that in this setting, the optimal fee is zero. This latter result questions, then, the conclusions we drew on optimal fee schedules based on the papers reviewed above where breadth was omitted as an instrument.

Hopenhayn and Mitchell (2001) suppose that ideas of type $\theta \in \Theta$ arrive to innovators with probability $g(\theta)$. This type is not observable by the planner. The development cost of the innovation is c. Innovators can make profits only if they obtain both length, T, and breadth, k, of protection so that profits are a function $\pi(k,T,\theta)$. The

[86] Pakes (1986) models patents as options in the sense that they are applied for at an exploratory stage, and information about their value is "revealed" over time. Early renewal decisions are based on both current (known) value and an option value to renew in the future. Higher value patents are those that are renewed longer, but this value is revealed over time, including an ever smaller option component and a greater certainty component. This concept of value is consistent with the models discussed here.

planner must maximise the social benefits of this protection regime, W, subject to the usual individual rationality and incentive compatibility constraints when it can set breadth, length of protection and also impose a fee. Hence, the problem is straightforward:

$$\max_{k(\theta),T(\theta),f(\theta)} \sum_{\Theta} W[k(\theta),T(\theta),\theta]g(\theta)$$

such that:

$\pi(k(\theta),T(\theta),\theta) - c - f(\theta) \geq 0$ (individual rationality)

$\pi(k(\theta),T(\theta),\theta) - c - f(\theta) \geq \pi(k(\hat{\theta}),t(\hat{\theta}),\theta) - c - f(\hat{\theta}) \forall \hat{\theta}$ (incentive compatibility)

$f(\theta) \geq 0$

In this setting, the fees are set optimally to zero under the appropriate sorting conditions. The reason is that the policy design problem optimises social welfare under constraints including covering research costs. Since fees raise these costs, they also tighten the constraint. This makes fees a relatively inefficient instrument compared to breadth or length of protection, which operate instead by raising value.

If innovations can be ranked according to how efficient length of protection is at generating surplus for the innovator, then the optimal contract involves a menu of length and breadth offered to different types of innovation where those innovations that get little value out of length get primarily breadth protection (large breadth and small length) while those that get large value out of length get primarily length protection (small breadth, large length). The justification is that some innovations are "fertile": they will generate follow-ons – developed by other firms -- that replace the original innovation in a short time span. As we have seen elsewhere, statutory length protection is of little value in this case since "effective length" will be much shorter. For example, suppose that the probability of arrival of a follow-on innovation in the first t periods after patenting for an innovation of fertility type θ is $p(t, \theta)$. The patentholder makes profits π per period and the basic innovation costs c to develop. Imagine that for some low fertility innovation type, θ_1, it is the case that an innovation with protection length T_1 and no breadth at all (so that all improvements are free to infringe) is just expected to generate enough discounted profits to cover the development cost. This may be the case because it is unlikely that improvements will be found during time span T_1. On the other hand, another innovation might have high fertility type, θ_2. Even with an infinitely-lived patent, it could be that this type might not be anticipated to cover its development cost with a zero breadth of protection. Instead, since improvements arrive very quickly, such an innovation requires positive breadth of protection, k: all improvements would be barred from the market for the duration of the patent. Hence, a patent system could involve protection levels (0, T_1) and (k, T_2), where $T_1 > T_2$. The low fertility type would strictly prefer the first type of protection and the high fertility type would strictly prefer the latter type. The patent authority can then screen innovations when it offers such a menu of protections.

In this framework, adding (leading) breadth is more effective at generating profits for high fertility innovations since this "slows down" the time at which the innovation will be replaced in a way that recalls ODonoghue's (1998) work. Notice that there is no licensing allowed here and that patentability and infringement are tightly linked. Further, length is adjusted in each case so that the "minimum" level of profits to induce innovation is always sent to the innovator in the optimal scheme. This begins to look like a very complex set of requirements to explain to the inventing public, let alone the public at large, and to implement at reasonable cost with realistic levers. That being said, some new proposals from industry have suggested having "deluxe" patents and "run of the mill" patents in a menu that would be at the disposal of patent applicants[87]. On the other hand, the point remains that while the Cornelli and Schankerman (1999) model appears practically implementable with existing tools, this mechanism – especially its conception and manipulation of breadth – would require considerable development to imagine in practice.

VI. How is the right secured? Enforcement

Intellectual property rights are only as good as their enforcement, and enforcement largely occurs by means of private suits brought by individual inventors or groups of inventors against other inventors. Indeed, Crampes and Langinier (2002) note that a patent merely grants the right to sue intruders that have been identified. Identification must be done by the patent holder at some monitoring cost, and even if infringers have been identified, the patent holder has the choice of how to react: by defending the patent in a court suit, settling out of court for some negotiated value or simply accommodating the entry. If the defence involves a counter suit questioning the validity of the patent, the patent holder could find that the upshot of litigation is to lose all rights to the intellectual property. Lemley and Shapiro (2005), developing an idea also put forward by Ayres and Klemperer (1999), propose modelling patents as "probabilistic" to capture the idea that patents only give a possibility – and not a guarantee – of a reward. The entry decision of a potential imitator depends on how aggressive the response to entry will be but also on the prior belief that the parties have about the likely strength of the patent, should it be challenged in court. The response to entry will, in turn, depend on the underlying patent strength, the characteristics of the firms involved, the market, and the cost of the various alternative strategies.

Lanjouw and Schankerman (2001) note that while the average litigation rate of patents is low – on the order of one percent of all patents – the probability of litigation of more valuable patents can be above 10% in some fields and more than 25% in pharmaceuticals. Infringement suits, such as those modelled by Crampes and Langinier, are more common than invalidity suits. Lanjouw and Schankerman note that due to the positive externality that a litigant creates by bringing an invalidity suit, their seeming paucity may not be surprising. In other words, a single litigant carries the entire burden of the trial costs, but if a patent is found to be invalid all potential users of the technology could benefit. Parties bringing private suits must weigh the high cost of the suits against the likelihood of covering these costs by extracting a high value in settlement. In terms of costs for those accused of infringement, Lanjouw and Lerner (2001) find that preliminary injunctions – bars on the allegedly infringing activity -- are used quite often (almost 20% of the cases they studied) as a remedy to infringing behaviour. An injunction can potentially "shut down" a

[87] IBM has proposed this two tier Community patent, see www.epip.eu/conferences/epip02/lectures/European%Interoperabily%20Patent%201.1.pdf.

business, which can run up the costs of being accused – rightly or wrongly-- very high.

The probabilistic nature of patents creates several effects. The first, following Ayres and Klemperer (1999), is that the probabilistic nature of patents induces a certain amount of infringement because infringement may pay off: the patent may not be upheld as valid (or may simply not be enforced). This compensates infringers for the risk they run of being held in violation of a valid patent. Their level of risk depends on the damages they must pay in the case of a finding of misconduct, and also the delay in the finding of infringement. A regime where patents are probabilistic but where disputes resolve slowly may benefit consumers without compromising innovation incentives[88] in the same way as compulsory licensing helped consumers in Tandon (1982) or restrictive competition policy helped in Gilbert and Shapiro (1990). Ayres and Klemperer suggest that an optimal regime would allow patent holders to choose from a menu of lengths of protection and probabilities of enforcement. Such a menu allows efficiency gains via lower deadweight loss when there is limited infringement, at the same time returning a target reward level to innovators, and harnessing the private information of patent-holders. In this sense, their ideas extend some of the insights of the Cornelli and Schankerman (1999) and Scotchmer (1999) frameworks to the case of probabilistic patents.

Ayres and Klemperer (1999) list a long set of caveats to their generally positive take on uncertain patent rights. This more negative side of the uncertainty coin has been developed by others in a series of papers[89]. One such negative is that the probabilistic nature of patents can create bias in the type of research that patents induce. The argument, put forward by Farrell and Shapiro (2007) is as follows. A standard response of an infringer is to challenge the validity of the patent at issue in court. One could think of an index of patent strength, *S*, where this index reflects the probability that the patent would withstand a test of its own validity (for example, a showing that that an invention fails to meet novelty, nonobviousness or usefulness criteria). Patents that are weak – have a low value of *S* -- can "punch above their weight" in terms of licensing revenues if private suits are not always brought.

As an example, consider an upstream lab that relies on licensing revenues from downstream industry for its income. A licensing scheme using per-unit royalties will be optimal for this patent holder as it allows the effective marginal cost of the licensees to be raised since the higher effective cost restrains downstream competition. Indeed, the monopoly outcome can be mimicked with the right choice of royalty. If there is also a fixed fee component to the licenses, then the fee can be used to distribute the profits from this newly "collusive" industry among the participants in this scheme, including the lab. Even a weak patent can be put to such a cynical purpose.

The antitrust status of the licensing agreement we have just outlined would be tenuous, of course. Even if we do not allow this type of collusive scheme, however, a

[88] It is important to have both uncertainty and delay to get full benefits. If disputes are resolved immediately, then with some probability the patentee sets the unconstrained monopoly price. If there is no uncertainty, but damages are set high enough to fully reimburse the patentee, then infringers cannot break even and so do not enter in the case where the patent is known to be "iron-clad".

[89] Farrell and Shapiro (2007), Lemley and Shapiro (2005, 2007), Shapiro (2008).

weak patent can generate surprisingly large revenues. As there is a positive externality for any single litigant to bringing a suit that reveals the true weakness of a patent, if a license for the patent is held by many firms, there is an under-incentive for any one to invest in the privately borne cost of litigation. For example, suppose a process innovation reduces marginal cost from *c* to *c*-*ε*. Let a potential licensee face the choice of joining up to a licensing agreement (now) or being excluded (forever) if she brings invalidity litigation. Additional litigants do not improve the chances that the patent will be found invalid in court. A potential licensee will accept the royalty rate proposed by the patent holder as long as it exceeds the payoff to litigation:

$$\pi(c-\varepsilon+r, c-\varepsilon+r) \geq S\pi(c, c-\varepsilon+r) + (1-S)\pi(c-\varepsilon, c-\varepsilon)$$

where profit, *π*, has two arguments: the effective marginal cost of the licensee and that of its rival. The left hand side of the inequality represents the profit earned if all firms accept the proposed royalty rate, *r*, on top of the lowered marginal cost of production that the innovation generates, *c*-*ε*. The first term on the right of the inequality is the payoff to being the one "excluded" firm that brings suit – and loses – weighted by the probability that this will occur. The last right hand term is the payoff to free access to the technology in the case that the suit is successful. Only one firm needs to bring the suit in order to generate this gain to all, so only a single firm – at most – will decide to be the "excluded" party. All the others would do better accepting the proposed royalty and awaiting the outcome of litigation. The maximum royalty rate that all firms will accept will depend on the importance to a firm of small changes in its own cost compared to small changes in the cost of the entire industry, as this determines how a change in royalty rate affects the right and left sides of this equation. In particular, if one thinks that the "fair" royalty rate for a patent of strength *S* that reduces marginal cost by amount *ε* is *Sε* there are many cases where the royalty will exceed this amount.

Hence, weak patents may be "overcompensated" relative to their true strength (as measured by benchmark *Sε*). Not only do we suffer deadweight loss because royalties are relatively high, but also research incentives are biased towards paths that result in "little steps" that would normally fall below patenting criteria because precisely these steps are those that are overcompensated in the market.

The policy implication of Farrell and Shapiro's model is that there is a benefit of patent review by the patent office to "weed out" weak patents before this licensing game ever occurs. The review would need to be directed at novelty and non-obviousness: if patents lack strength due to incorporating subject matter that has not heretofore been patented, it may be precisely the more creative and big steps that result in "weak patents". Hence, the policy recommendation for the review must place a specific interpretation on "weakness". Encoua and Lefouili (2008) note that the argument we have made shows bias for small innovations. If one allows larger – drastic – innovations in a similar framework, they can generate the result that weak patents punch *below* their weight under the right assumptions[90]. Despite this, the point made by Farrell and Shapiro is valid in that correcting "errors" in awarding patents via the courts may create distortions in some cases that could be avoided by more careful patent office review. While totally eliminating those errors is probably

[90] See Encoua and Lefouili (2008) for derivations.

too costly, as Lemley (2001) suggests, simply relying on private suits to sort out errors also carries significant costs that could be avoided by patent office review[91]. A compromise that eliminates more errors would probably be desirable.

Hence, one "take" on modelling enforcement issues is to model patents as generating a probability of benefits, but not a guarantee. The outcome of the probabilistic models depends, of course, on what remedies are allowed to the patent holder in the case of a successful finding of infringement. This leads to modelling that compares damage systems to injunction systems as means of upholding the right. Indeed, a fundamental legal issue is whether market conduct should be managed by means of completely suppressing certain practices, for example by enjoining behaviour, or by allowing these practices subject to the payment of damages in the case of harm. Hylton (2006), commenting on an argument made earlier by Calabresi and Melamed (1972), notes that property rules – such as the intellectual property right we have analysed so far -- prohibit others from infringing the property right without first gaining consent. In contrast, liability rules do not require consent, but rather simply require the payment of damages when loss has occurred. Whereas liability may simply make an activity unprofitable, an injunction can directly prevent conduct that infringes that right[92]. Formalising this idea, Anton and Yao (2004), discussed earlier, explicitly includes the possibility of damages as a "reward" that inventors can collect in the case of imitation that is triggered by the patent disclosure. In such a system, triggering imitation can be good for the patent holder, if we assume that damages are not too expensive to collect and are awarded based on the true value of the innovation. Indeed, in that model damages and licenses would serve some of the same functions. Hylton focuses more narrowly on the cost side, examining the role of transaction costs and the distribution of valuations of the property to determine whether property rules or liability rules are more desirable. For example, if bargaining is not possible, then a high value user might effectively be barred from creating value under property rules. This case would be equivalent to the case of no licensing in the earlier papers we reviewed on patent design. Damages also have their own problems: asymmetric information on the magnitude of the damages could create a wedge between the actual damage award granted by the court and the damage truly incurred by the property holder. Hence, the nature of market failures in the transfer of intellectual property can induce a ranking of liability and property rule systems.

Ayres and Klemperer (1999) focus largely on damages, arguing that these will promote limited infringement as imitators "try their luck" against a patent holder who may not (successfully) enforce her patent. Limited infringement may be just what an efficient system should aim for, as limited infringement effectively puts a cap on deadweight loss, but also allows per-period rewards to research activity to accumulate. In this sense, damages have some of the benefits that a royalty cap did in Tandon's compulsory licensing scheme, which we discussed in the single innovation

[91] On the other hand, uncertainty about which of two competing innovators' patent claims might be valid might result in a large spread of small expected rewards, inducing a large amount of entry or more efficient bargaining to the extent that this results in "joint ownership" of the right to produce. See footnote 1 and references contained in Ayres and Klemperer (1999).

[92] Ayres and Klemperer draw an analogy to type I and type II errors: Let the null hypothesis be that the patent is valid. A damage system corresponds to a system where a valid patent might not be enforced, creating a type I error. An injunctions system corresponds to one where an invalid patent might be enforced, creating a type II error.

model section. Injunctions do not have this benefit. They amount to blanket prohibitions and so cannot achieve the same efficiency gains of limited infringement: either infringement is unlimited or it does not happen at all[93].

Schankerman and Scotchmer (2001) make the bargaining role of damages and injunctions more precise to compare their effects. They show that different rules allow for different credible threats in the case that negotiations for access to a patent break down. These different levels of credible threats can affect the division of profits between a violator and a victim. For example, suppose that an independent firm has developed a profitable application of a patented innovation that infringes the original patent. If the first patent holder has the ability to enjoin an infringing firm, then a violator has a "credible threat" *not* to infringe. If a violator to credibly "refuse" to create value, the threat of refusal can be used to extract value from the original patent holder in an access agreement. Instead, suppose that an inventor facing infringement by a high-value violator could use damages to collect this value *ex post*. Infringement is good for the inventor if the infringer can use the invention to create more value than the inventor could. This value can simply be extracted *ex post* through a damage settlement. Hence, the ability of damages to allow for value creating activity while distributing the gains back to the original patent holder can be more valuable to the inventor than the ability to enjoin an infringing firm. This sort of role for damages is very much in the spirit of the socially valuable role that licensing performed in Scotchmer's earlier work on cumulative innovation, summarised above. Hylton's addition to this is to caution that licensing markets may not work well, and courts may also not have the information they need to award appropriate damages[94].

A final set of papers on enforcement pairs litigation issues with the case of complementary innovation ("patent thickets"). One of the basic forces in models with both these considerations is illustrated by Lemley and Shapiro (2007). This paper investigates the effect of injunctions on bargaining for (licensing) access to patents. Similar to the benchmarking exercise conducted by Farrell and Shapiro, above, Lemley and Shapiro propose a benchmark "fair" level of compensation for any one patent that reflects the value contributed by that patent to the product, a measure of bargaining skill in negotiations, and the strength of the patent. It then compares the actual compensation received in a bargaining game to this benchmark level. Hence, the exercise is similar to Farrell and Shapiro (2007), but is in the context of "complex" products composed of a large number of patented elements. Here, the benchmark is more complex as well.

Lemley and Shapiro find that, where a product reads on a large number of patents, the negotiated royalty rate can exceed this benchmark level. The reason is based on hold-up. Suppose that the cost of developing the product is largely sunk at the point when the inventor and the producer attempt to negotiate a royalty. In this case, the holder of

---

[93] Maurer and Scotchmer (1999) use a similar argument for the benefits of limited entry to suggest that an independent invention defence should be allowed in cases of infringement.

[94] Boyce and Hollis (2007) make an even stronger point against injunctions, suggesting that the way they are implemented creates the possibility that they can be used as a "court-ordered collusive scheme" since they protect the patent monopoly without compensating consumers. Furthermore, they claim that US patent law – perversely -- can allow patent holders to gain more from the injunction process if the imitator is found not to have infringed. This creates an incentive to seek a preliminary injunction when patents are weakest.

a single patent can "hold up" the producer for the entire value of the product if royalty negotiations break down when the alternative to a license is an injunction against the product's sale. This potentially gives the holder of a single patent the power to extract much more than its own contribution to the final value of the product. Indeed, if a product reads on many small patents, each individual patent holder may be very keen to exploit this form of hold up to raise its licensing revenues. Hall and Ziedonis (2001) document such hold up in the face of large sunk investments[95].

This type of hold-up can have several undesirable effects. First, if patent holders can extract very large royalties by negotiating independently with the producer, it can result in the patent holders' not wishing to join standard setting groups that might license patents in a package. While the patent holders could potentially increase the total surplus to be shared by pooling together their patents, a coordinated solution is unlikely to arise given that each patent holder wishes to exploit hold up to her own advantage. Further, negotiating licenses individually for a large number of patents could also be an extremely costly process, which could create direct social welfare losses. Secondly, producers might avoid entering an industry where this form of hold-up might arise, causing under-provision of certain types of product and further attendant welfare losses. Of course, designing around such patents may be the best way to avoid these overcharges, but if re-design is costly it also causes losses in its turn. The paper concludes that various policy solutions, including limiting the use of injunctions, and some imposition of reasonable royalty calculations, could go a long way to resolving these problems. While this approach hammers home a point that Scotchmer brought up earlier to wit, *ex ante* licensing works better than *ex post* licensing to efficiently coordinate technology sharing in the case of externalities, the presumption of this work is that *ex post* licensing is the more relevant case. After all, it might be in the interests of the "troll" to *wait* until investments have been sunk to make his presence known to potential victims. In other words, the troll hides under the bridge and then emerges to extract large fees. On the other hand, Siebert and von Graevenitz (2005) find evidence for increased *ex ante* licensing activity in the semiconductor industry, where patent thickets are generally thought to be present[96].

These arguments rely on an assumption of complementarity and, as Lerner and Tirole have pointed out, it may be quite difficult to determine how complementary or substitutable they are and this may change[97]. Indeed, Galasso and Schankerman (2008) analyse the effect of fragmentation of patent claims to a valuable "reward" on legal disputes when patents may be less than perfect complements. Imagine that a single product that generates market value reads on a large number of patents. If greater fragmentation reduces the contribution of any single patent to the final product – i.e., if patents are not perfect complements – the value of litigating (or continuing to litigate) any single patent falls because the expected damages fall[98]. They show

[95] For related work, see also Gilbert and Katz (2006) and Gilbert and Katz (2007). The term "patent troll" has been coined to refer to entities that aggressively enforce patents in order to extract exorbitant licensing fees by exploiting hold-up. A series of actions involving NTP, Inc. and Research in Motion, maker of the BlackBerry, has been held up as a canonical example of patent trolling. In that case, it was claimed that the "troll" was in hiding until the value of the final product was created.
[96] Geradin et al (2007) evaluate proposals for testing *ex ante* promises to license at reasonable and non-discriminatory terms.
[97] Choi (2003) incorporates litigation into a patent pool model, finding that pool members with weak patents tend not to bring suit against each other because suits provoke counter-suits.
[98] See also Lichtman (2006) on this point.

empirically that greater fragmentation is associated with *less* delay *per dispute* in settling patent litigation, which suggests that each litigated patent must have smaller value. If we were in a world where fragmentation involved strictly complementary patents, this should not be the case since each patent is vital to the final product and so has equal "value". Hence, they reason that the sort of hold-up effects we saw in the Lemley and Shapiro work are being dominated empirically by the effect of decreased significance of any single patent.

In a separate point, Galasso and Schankerman (2008) note that, to the extent that the Unified Court of Appeals has introduced less uncertainty about the outcome of court disputes, this decreases the role for information asymmetries in the bargaining process that spawns litigation. In other words, whether the court is or is not biased in its judgements is less relevant to the speed of settlement than the fact that the outcome may be more *certain* with a court that always rules a particular way. As a result, patent litigation should be observed to progress more quickly to settlement under the Unified Court of Appeals. This is what they find, but the finding runs counter to some of the received wisdom on the events of recent years that litigation has been "excessive" in areas where rights are fragmented. They reconcile these views by suggesting that even though they find that delay has fallen per dispute, the total settlement time could have increased because the total number of disputes would be expected to rise with the degree of fragmentation of rights. In other words, if technologies are more "complex" nowadays there are more points of conflict, so the benchmark for the "expected" amount of litigation should rise with fragmentation.

Summarising, one way of modelling the role of enforcement is to say that patents are not "iron clad" rights, but instead are probabilistic. Here, a main finding is that even patents that are unlikely to withstand a court challenge could generate very large licensing revenues because the incentive for any one user to challenge the patent could be socially too small. This could argue in the US for instating post-patent review along the lines of the European system so as to reduce the burden of weak patents. A second tack to take is to examine the tools at the disposal of a patent holder in the case of infringement and see how varying this set of tools affects efficiency. A role of the tool is to set the threat that can be held out in case negotiations for access to patent rights break down. Liability rules that simply assign damages in the case of harm from infringement can act to compensate innovators for use of their innovations while allocating the innovation to the high value users. Some have argued for injunctions on the basis of courts' inability to award damages that reflect value to innovations' creators. However, injunctions combined with hold-up can allow owners of individual patents to extract such high royalty rates that producers who must combine patents to create valuable end products could be put off entering in the first place. This argues for damages to be used instead as a remedy for infringement.

VII. Patent Rights and Competition Policy

Competition law and intellectual property rights have developed as two separate areas of law, each with its own goals and methods. While not a focus of the review, it should be clear from the preceding sections that competition policy and patent policy interact jointly to determine innovation incentives. Recalling Gilbert and Shapiro's (1990) perspective, competition policy can determine the flow rewards to patenting for any given patent right policy. Competition policy serves this function by determining the parameters for use*,* while patent rights determine the parameters for exclusion.

While one might think that a limit on flow rewards might necessarily contradict the goals of patent policy, Ayres and Klemperer (1999) put forward the argument that competition policy and reward to innovators need not be "at odds". As we saw in the review of single innovation models, small restrictions on the patent-holder's pricing from the unconstrained monopoly level will have a second-order effect on the monopolist's profits (since profits were maximised at the unconstrained price) but a first order effect on deadweight loss (since social welfare was not)[99]. By re-balancing the patent right and the parameters of use to yield a longer but lower level of market power, one could create efficiency gains that would at once benefit consumers and not hurt inventors. To implement this idea some have used Kaplow's (1984) ratio test: the ratio of the patentee's incremental reward to the incremental social loss of a given practice. This ratio is a litmus test for allowing a practice: the higher the ratio, the more desirable the practice. Ayres and Klemperer's point is that under such a "ratio test", allowing unconstrained monopoly power probably looks like a bad idea: the ratio is zero at the unconstrained monopoly point.

Evans and Schmalensee (2002) point out that innovators often compete *for* markets, rather than compete on the margin *in* a market, as in Ayres and Klemperer's treatment. Competition policy issues for innovative industries may centre less controlling per period reward for a given market than on practices that tend to exclude other competitors and their innovations from emerging markets. This gains particular force in network industries, of course. A third area where competition policy can interact with innovation incentives is that it can define when and how coordination can occur amongst independent players in an industry. As such, it affects the scope for licensing agreements, which we have seen plays a crucial role in patent design for cumulative and complementary innovation.

A further difficulty is that, if we think of competition policy as regulating structural characteristics of markets, it is not at all clear what sort of market structure promotes innovation best. In general, a firm that already is earning high profits in an industry will have little incentive to "replace" itself by innovating more. On the other hand, this sort of firm might wish to innovate in order to prevent another firm from taking over the market from outside. The effect of product market concentration on innovation is unclear, then. Indeed, which product market structure is socially best in terms of innovation incentives depends on the interplay of various factors that must be balanced quite specifically in order to generate any clear guidance[100]. Further, the structure of the innovative process itself determines whether there is socially too much or too little incentive for innovation in the economy in the first place[101]. As Shapiro (2007) points out, it is not clear whether we would align social and private incentives to innovate by increasing or decreasing innovation incentives. If we don't know what market structure to aim for or whether we are even above or below "target" as a starting point for analysing policy, attempting to give any general guidance on structure can be daunting.

---

[99] For demand that is not too convex, this will hold for larger reductions in price as well.

[100] See Aghion *et al* (2005), for example.

[101] See Gilbert (1995) for a discussion of "innovation markets", how they are distinguished from product markets and how their structure is taken into account in antitrust policy.

Regibeau and Rockett (2007) suggest some overall rules for the interface between competition policy and intellectual property[102]. Their starting point is that the exclusionary rights granted by intellectual property law do not necessarily confer monopoly rents, but they can only be effective at stimulating innovation if they sometimes do. Indeed, as many of the models we have reviewed have shown, any single innovation may either fail to find a profitable market or may be pre-empted so rapidly by further advances that it generates very little value for the innovator. Still, it is the expectation of rents *ex ante* that generates innovation in the first place. Hence, some level of rents must be made available. In point of fact, competition policy only intervenes selectively after an asset – of any type – becomes the basis for significant monopoly rents. A consequence of this difference in timing is that – even if the goals and skills of intellectual rights policy and competition policy authorities were the same -- the information available to each of these authorities can differ. Indeed, to the extent that competition authorities tend to have access to more detailed information than was available at the time property rights were granted, there is a great temptation to revisit the trade-off between innovation incentives and the deadweight loss resulting from intellectual property rights. Another consequence is that the intervention does not fall evenly across all property, which can feed back on the level of expected reward.

If one takes the perspective that intellectual property is there to generate a certain reward for innovators then, as Ayres and Klemperer (1999) and Maurer and Scotchmer (1999) have emphasised, that reward can be generated even if competition law limits the extent to which a rights holder can exercise the property right. In this sense, once competition law is "fixed" property rights design can be adjusted to accommodate it and still achieve its basic goals. The only requirement is that competition policy not completely expropriate those gains. Second, as the *expectation* of gains is what counts, intellectual property law need not adjust to *individual* competition law decisions except to the extent that they represent a change in competition law policy. On the other side of the balance, any systematic attempt to revisit the trade-off between static and dynamic efficiency by competition law could be undesirable. These trade-offs – evaluated from an *ex ante* viewpoint -- are what determine intellectual property right design in the first place, so revisiting these trade-offs later – from an *ex post* viewpoint -- are unlikely to improve innovation incentives (which are determined by *ex ante* evaluations). Indeed, there is a real risk of regulatory opportunism at later stages. Given that *ex post* the socially optimal policy is free access to intellectual property, the incentives of authorities intervening selectively *after* the property has been created are not aligned with those of an authority that wishes to optimise *ex ante* innovation incentives. This is particularly salient at the level of an individual case: to the extent that each case affects expectations of reward very little, it is unclear what there is to "lose" by expropriating rights in a single case. Hence, there is an argument for commitment, whatever policy is chosen.

Another way of thinking of this interaction is to advocate policy made in a coordinated way, rather than delegated to individual court decisions, a point stressed

[102] See also Anderson and Gallini (1998) for a volume covering many aspects of intellectual property law and competition policy.

by Scotchmer (2002). As we have argued, individual courts may not have the incentives to take decisions in a way that preserves *ex ante* innovation incentives. In this sense, it is well to make policy at the level of the legislature rather than in the courts.

A final general issue is clarity. If antitrust policy is confusing or contradictory in its treatment of certain practices, it can be difficult for inventors to calculate any particular reward for investment.

This set of principles does not give much guidance on a more specific level of particular competition laws and particular elements of intellectual property policy. While the argument in the previous paragraphs applies quite generally – even beyond intellectual property to other types of property that can be the basis of monopoly power – there are interactions between these two bodies of law that are quite specific to intellectual property and must be modelled specifically. Licensing policy is an area that has received a lot of attention, where the concern is to preserve the efficiency enhancing aspects of licenses while preventing such undesirable elements as over-pricing. Maurer and Scotchmer (2006) recommend that a reasonable policy objective should be that a licensor should be able to earn as much profit by licensing as by producing the product herself in an equivalent production environment. The idea here is not to penalise licensing activity that may, for example, be necessary because of financial constraints faced by the inventor. This principle of "profit neutrality" can give a useful benchmark to authorities generating specific rules towards licensing activity and fees.

Shapiro (2003) examines another type of "neutrality" to address antitrust treatment of proposed settlements of patent disputes. He suggests that these should generate at least as much consumers' surplus as they would have accrued if litigation had ensued. The exact nature of the settlement does not need to be specified: it could be merger, joint venture, licensing or other agreements. The point is that while the invention incentives need to be preserved, efficiency is a concern as well. We need a precise rule to balance the two. Formally, the problem of settlement design is the "dual" of a Ramsey pricing problem where we solve:

$$\underset{x}{Max}\, \pi(x) \text{ subject to: } s(x) \geq \bar{s}$$

where x is a set of actions that the two parties can specify in a contract *ex ante*, π is industry profits, s is consumers' surplus, and $\bar{s}$ is the surplus that would result from litigation. Given that litigation remains an alternative, the parties can always do better to settle on these terms, even without consideration of the expense of a trial.

What remains is to translate this general rule into settlement policies. For example, in the case of licensing, we could propose capping the royalty rate so as to guarantee a minimum level of surplus to consumers. In the case of mergers, we could apply a factor -- taking into account that a merger involves intellectual property issues -- to the normal offsetting efficiencies we might demand to justify a merger's anticompetitive harm. For example, define a ratio[103] of the harm caused to consumers by the merger with patent issues over the harm that would be caused by a merger with no patent issues. If the standard merger would require offsetting efficiencies, *E*, to benefit consumers, a merger involving patenting parties would require *E* times that ratio to be justified. Hence, we have a "scaling factor" to apply to standard merger guidelines and procedures, where the scaling depends on patent strength[104].

More specific issues have been the subject of a range of models incorporating both competition and intellectual property policy. The Shapiro (2007) benchmark licensing pricing policy in the context of multiple (complementary) innovations can be used to evaluate whether a license is being issued at fair and reasonable terms[105]. Licensing restrictions such as grantbacks (Choi (2002), Van Dijk (2000)), cross licensing (Fershtman and Kamien (1992), Eswaran (1994), Choi (2003)), licensing for standard setting (Farrell et al (2007)), as well as joint ventures (Scotchmer (1998)) and tying (Whinston (1990, 2001), have been investigated[106]. An area where many of the conflicts have emerged has been in industries with network effects, so special attention has been paid to modelling in this area (see Farrell and Katz, (1998, 2000), for example). Most of these models rely on specific structural features to generate their results or require judgments that may be difficult in practice (such whether patents are substitutable or complementary). Hence, it is difficult to make watertight rules based on them that policy-makers can easily implement in a wide variety of cases. Developing rules based on a consensus in the literature may need to wait until more modelling is done so that a considered consensus can emerge. Alternatively, advocating a rule of reason approach under some broad guidelines may be about as far as theory can go.

Finally, in contrast to most of the papers outlined here, the interaction between competition policy and innovation has been treated in the growth context to give slightly different conclusions. In a recent example of this, Segal and Whinston (2008)

---

[103] Formally, the ratio equals $\frac{\bar{s} - s_M}{s_D - s_M}$, where s_D is consumers' surplus in the case of duopoly and s_M is surplus in the case of monopoly. To see how this works in practice, take two polar cases of patent strength. If a patent is iron-clad and if the resolution of litigation would be immediate the "reservation" level of consumers' surplus is just the monopoly level and merger creates no further harm to consumers when patents are present than without. Indeed, the numerator of the ratio is zero, so no efficiency gains are necessary to justify the merger. If the patent has no chance of surviving the court challenge, on the other hand, the outcome of litigation is the duopoly payoff. In this case, the ratio is one since both the numerator and denominator are the difference between duopoly and monopoly surplus. Here, the efficiency gains that would be required neglecting patent issues would be the same as those required to justify the merger in the presence of patents. See Shapiro (2003) for a full discussion.

[104] As a practical issue, both the solution to the maximisation problem and the level of the ratio depend on the strength of the patent, as this determines the outcome to litigation. Shapiro notes that the strength could be difficult for any of the parties, let alone an outside authority, to evaluate accurately. Katz and Shelanski (2006, 2007) and Carlton and Gerntner (2003) weigh other factors, such as the reduction in duplicative effort, the benefits of specialisation, and various spillover benefits against the reduction in competition from merger when comparing merger in innovative industries to "standard" cases.

[105] Also see Gilbert (1995) and Gilbert and Sunshine (1995) on the Department of Justice 1995 guidelines and provisions involving licensing of complementary factors.

[106] See also Gilbert and Shapiro (1997), who map out a large number of issues that are developed in later work by both authors.

note that competition policies that protect new entrants from exclusionary behaviour by incumbents can raise entrant profits, thereby encouraging innovation by entrants. On the other hand these same policies will come back to haunt the entrants in the future, should they be successful. This future lower profit to incumbency can eventually slow the rate of innovation. Hence, an antitrust policy that favours new entrants affects the time distribution of benefits from innovation, "front loading" those benefits. This sort of timing issue is something that a growth model can highlight particularly well. Segal and Whinston analyse the tension between rewarding entrants and incumbents in the light of growth goals and several specific antitrust policies, finding that in some cases, policies benefit both entrants and incumbents while in others there is a conflict of interest between the two parties.

VIII. Conclusions

This is far from a settled literature. Many basic points remain unresolved, such as what style of protection should be used to promote innovation, and the level of incentives compared to the social optimum. Economic modelling of the complex features of the administrative process of getting a patent, the litigation process, and the interaction between specific areas of intellectual property and competition law are ample, but still do not cover the range of issues actually faced by the wide range of actors involved in getting an innovation out to market. Some topics are areas where feeling runs high. A quick search of the web, for example, indicates that "patent trolls" can be viewed as heroes or as demons, with a wide number of arguments on either side that have only been partially modelled in the academic literature. Finally, innovation policy is an area where there is intense interest on the part of policy makers. If one wishes to make a contribution to a debate that is viewed as pressing, this is a good area to enter.

There are many issues that have not been treated much in this chapter -- since they have been developed relatively less in existing work – but which are nonetheless very important. One such issue is how to compensate and motivate scientists to conduct innovative activity. Aghion and Tirole (1994) consider the question of whether research units (such as the scientists themselves) or customers (such as the manufacturing firm that might use the patent to produce value) should own patents: the issue is not so much the design of the property right as the allocation of it across interested parties when financing, creating value, and rewards potentially result from the effort of a variety of differing agents. These authors note that the magnitude of research activity as well as its nature (such as the size of innovation) can change with different organisational structures. While some work has followed on scientists, their contracts and incentives, and the organisations in which they work[107], a lot still remains to be done.

This survey has been quite narrow: it has not covered all of intellectual property. Copyright and trademark law have been excluded – with regret. Many of the issues are similar between copyright and patent, but not all. Indeed, technology is forging

[107] Kim and Marschke (2005) study labour mobility issues and Sena (2004) reviews spillovers and cooperative agreements. Sorenson and Fleming (2004) relate norms and institutions of science and their effect on patenting activity. Severinov (2001) studies confidentiality agreements.

ahead so rapidly in the area of digital media that copyright issues are a fertile area of work. For example, an issue that comes up with copyrighted digital media but less with patented material is whether to protect the intellectual property with legal rights or with technical boundaries (such as encryption) [108]. The economics of trademarks is rather distinct from that of patents and copyright, involving reputation mechanisms, consumer search and repeat purchase. Still, to the extent that many products employ many methods of protection all at once-- patenting some aspects, employing secrecy for others, copyrighting design aspects and trademarking the product as a whole -- investigating these rights as they are used together could be a fruitful area to pursue as well. Indeed, using these three tools together to create the sort of exclusion zone around a product that is assumed in some of the single-innovation models is a goal of such multiple coverage but little work outside of the interaction of trade secrecy and patents has been done[109].

A third area where there is relatively little work is the exploration of alternative instruments and institutions that are currently available to influence innovative activity. For example, Acharya and Subramanian (2007) examine the effect of bankruptcy codes on innovation, where innovation is a relatively "risky" investment. Financial structure affects innovation, so that the incentive to innovate depends on how creditor-friendly or how debtor-friendly bankruptcy codes are. Given that there appears to be an empirical link between financial structure and innovation (see Hall (2002) and references therein), a more systematic investigation of financial regulations' effect on innovation could be useful.

Finally, some recent work has examined the effect of local versus global diffusion of information on behaviour (for example, see Boncinelli (2008)). While one could imagine that information about innovation might travel easily amongst members of a focussed set of researchers by word of mouth or on the conference circuit, it is possible that the function of patents as repositories of information might help to spread information to those who would not normally receive it. This could affect the trajectory of future developments as well as their speed. Developments in the field of social network theory could be helpful to diagnose the effectiveness of patent repositories versus the use of other tools targeted more at "key" individuals.

---

[108] See, for example, Menell and Scotchmer (2007) for a discussion of the law and economics of trademark and copyright. Varian (2005) provides a recent review and opinion on copyright.
[109] For a discussion from the legal perspective, see Merges et al. (2006). The interaction trademarks and patents is studied in Manelli et al. (1994).

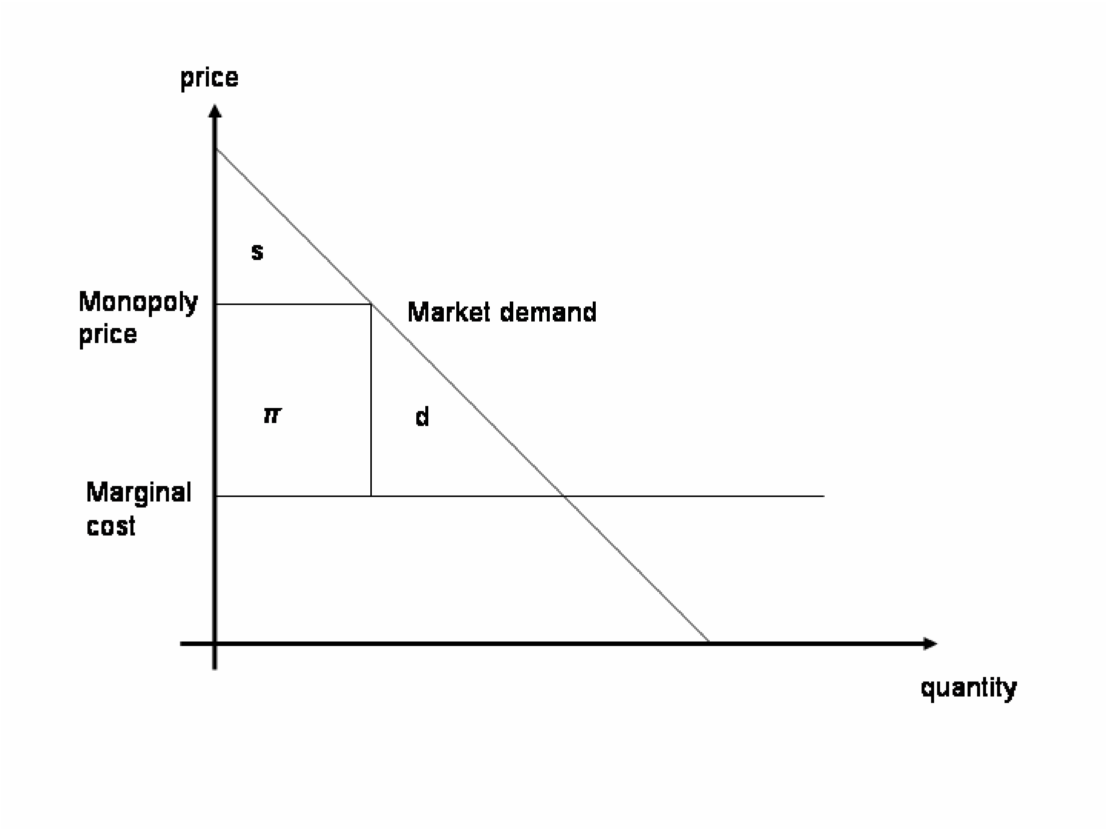
price
s
Monopoly
price
Market demand
π
d
Marginal
cost
quantity

Figure 1

References

Acharya, V., Subramanian, K. (2007) "Bankruptcy Codes and Innovation". Centre for Economic Policy Research Discussion Paper 6307.

Aghion, P., Tirole, J. (1994) "The Management of Innovation". The Quarterly Journal of Economics 109(4), 1185-1209.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., Howitt, P. (2005) "Competition and Innovation: An Inverted-U Relationship". Quarterly Journal of Economics 120(2), 701-728.

Allison, J., Lemley, M. (1998) "Empirical Analysis of the Validity of Litigated Patents". American Intellectual Property Law Association Quarterly Journal 26(3), 185-275.

Anderson, R., Gallini, N., Eds (1998) Competition Policy and Intellectual Property Rights in the Knowledge-Based Economy. University of Calgary Press: Calgary.

Anton, J., Yao, D. (1994) "Expropriation and Inventions: Appropriable Rents in the Absence of Property Rights". The American Economic Review, 84(1), 190-209.

Anton, J., Yao, D. (2002) "The Sale of Ideas: Strategic Disclosure, Property Rights and Contracting". Review of Economics Studies 69(3), 513-531.

Anton, J., Yao, D., (2003) "Patents, Invalidity, and the Strategic Transmission of Enabling Information". Journal of Economics & Management Strategy 12(2), 151-178.

Anton, J., Yao, D. (2004) "Little Patents and Big Secrets: Managing Intellectual Property". The RAND Journal of Economics 35(1), 1-122.

Anton, J., Greene, H., Yao, D. "Policy Implications of Weak Patent Rights". In A. Jaffe, J. Lerner and S. Stern, Eds. Innovation Policy and the Economy 6(1). MIT Press: Cambridge, Ma., 1-26.

Aoki, R., Spiegel, Y. (1998) "Public Disclosure of Patent Applications, R&D and Welfare". Foerder Institute Working Paper 30-98.

Arora, A., Fosfuri, A., Gambardella, A. (2001) Markets for Technology: The Economics of Innovation and Corporate Strategy. MIT Press: Cambridge, MA.

Arora, A., Ceccagnoli, M., Cohen, W. (2003) "R&D and the Patent Premium". National Bureau of Economic Research Working Paper 9431.

Arora, A., Merges, R. (2004) Specialized Supply Firms, Property Rights, and Firm Boundaries. Industrial and Corporate Change 13, 451-475.

Arrow, K. (1962) “Economic Welfare and the Allocation of Resources to Invention”. In Nelson, R., Ed. The Rate and Direction of Inventive Activity: Economic and Social Factors. National Bureau of Economic Research: Cambridge, MA., 609-626.

Ayres, I., Klemperer, P. (1999) “Limiting Patentees’ Market Power without Reducing Innovation Incentives: The Perverse Benefits of Uncertainty and Non-Injunctive Remedies”. Michigan Law Review 97(2), 985-1033.

Baker, S., Mezzetti, C. (2005) “Disclosure as a Strategy in the Patent Race”. Journal of Law and Economics 48(1), 173-194.

Bar, T. (2006) “Defensive Publications in an R&D Race”. Journal of Economics & Management Strategy 15(1), 229-254.

Belenzon, S. (2006) “Knowledge Flow and Sequential innovation: Implications for Technology Diffusion, R&D and Market Value. Oxford University Working Paper 259.

Bessen, J. (2006) “The Value of U.S. Patents by Owner and Patent Characteristics” Boston University School of Law Working Paper 06-46.

Bessen, J., Maskin, (2007) “Sequential Innovation, Patents and Innovation”. The RAND Journal of Economics, *forthcoming.*

Bessen, J., Meurer, M. (2008a) “The Private Costs of Patent Litigation”. Boston University School of Law Working Paper 07-08.

Bessen, J., Meurer, M. (2008b) “The Value of U.S. patents by Owner and Patent Characteristics”. Research Policy. *forthcoming*.

Bidgoli, H. (2009) The Handbook of Technology Management, Vol. III: Management Support Systems, Electronic Commerce, Legal and Security Considerations. J. Wiley & Sons: New Jersey. *forthcoming*.

Bloom, N., Schankerman, M., Van Reenen, J. (2007) “Identifying Technology Spillovers and Product Market Rivalry”. National Bureau of Economic Research Working Paper 13060, April.

Boldrin, M., Levine, D. (2003) “Rent-seeking and Innovation”. Journal of Monetary Economics 51(1), 127-160.

Boncinelli, L. (2008) “Global vs. Local Information in (Anti-)Coordination Problems with Imitators”. The B.E. Journal of Theoretical Economics 8(1) (Topics) Article 17.

Boyce, J., Hollis, A. (2007) “Preliminary Injunctions and Damage Rules in Patent Law”. Journal of Economics and Management Strategy 16(2), 385-405.

Brunt, l, Lerner, J. Nicholas, T. “Inducement Prizes and Innovation”. Centre for Economic Policy research Discussion Paper 6917.

Calabresi, G., Melamed, D. (1972) "Property Rules, Liability Rules and Inalienability: One View of the Cathedral". Harvard Law Review 85(6), 1089-1128.

Carlton, D., Gertner, R. (2003) "Intellectual Property, Antitrust, and Strategic Behavior" in A. Jaffe, J. Lerner and S. Stern, Eds. Innovation Policy and the Economy (1). MIT Press: Cambridge, Ma., 29-59.

Cassiman, B., Veugelers, R. (2002) "R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium". American Economic Review 92(4), 1169-1184.

Chen, M., Iyigun, M. (2006) "Patent Protection and Strategic Delays in Technology Development: Implications for Economic Growth". University of Colorado Working Paper 04-09, revised 2007.

Choi, J. (2002) "A Dynamic Analysis of Licensing: The 'Boomerang' Effect and Grant-Back Clauses". International Economic Review 43(3), 803-829.

Choi, J. (2003) "Patent Pools and Cross-Licensing in the Shadow of Patent Litigation". CESifo Working Paper 1070 (2003).

Cohen, W., Levinthal, D. (1989) "Innovation and Learning: The Two Faces of R&D". The Economic Journal 99(397), 569-596.

Cohen, W. Nelson, R. Walsh (2000) "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)". National Bureau of Economic Research Working Paper 7552.

Comino, S., Manenti, F., Nicolo, A. (2007) "Sequential Innovations with Unobservable Follow-on Investments" Universita Degli Studi Di Padova "Marco Fanno" Working Paper 41.

Cornelli, F., Schankerman, M. (1999) "Patent Renewals and R&D Incentives". The RAND Journal of Economics 30(2), 197-213.

Crampes, C., Langinier, C.. (2002) "Litigation and Settlement in Patent Infringement Cases". The RAND Journal of Economics 33(4), 258-274.

Denicolo, V. (1996) "Patent Races and Optimal Patent Breadth and Length". The Journal of Industrial Economics 44(3), 249-265.

Denicolo, V. (2000) "Two-Stage Patent Races and Patent Policy". The RAND Journal of Economics 31(3), 488-501.

Denicolo, V., Franzoni, L. (2004) "The Contract Theory of Patents". International Review of Law and Economics 23(4), 365-380.

Dewatripont, M., Legros, P. (2008) "'Essential' Patents, FRAND Royalties and Technological Standards". Centre for Economic Policy Research Discussion Paper 6925.

Ebert, L. (2004) "How High are the Grant Rates at the USPTO?" Journal of the Patent and Trademark Office Society 86(7), 568-576.

Eisenberg, R. (1989) "Patents and the Progress of Science: Exclusive Rights and Experimental Use". The University of Chicago Law Review 56, 1017-1086.

Encoua, D., Lefouili, Y. (2008) "Licensing Uncertain Patents: Per-Unit Royalty vs. Up-Front Fee". Paris School of Economics, University of Paris-I Pantheon Sorbonne, mimeo.

Eswaran, M. (1994) "Cross-Licensing of Competing Patents and a Facilitating Device". Canadian Journal of Economics 28(3), 698-708.

Evans, D., Schmalensee, R. (2002) "Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries". In A. Jaffe, J. Lerner and S. Stern, Eds. Innovation Policy and the Economy 2, 1-49.

Farrell (1995) "Some Arguments for Weaker Intellectual Property Protection in Network Industries". StandardView 3(2), 46-49.

Farrell, J., Katz, M. (1998) "The Effects of Antitrust and Intellectual Property Law on Compatibility and Innovation". Antitrust Law Bulletin 43(3), 609-650.

Farrell, J., Katz, M. (2000) "Innovation, Rent Extraction, and Integration in System Markets". Journal of Industrial Economics 48(4), 413-432.

Farrell, J., Shapiro, C. (2007) "How Strong are Weak Patents?". Competition Policy Center Paper CPC05-54.

Farrell, J., Hayes, J., Shapiro, C., Sullivan, T. (2007) "Standard Setting, Patents, and Hold-Up". University of California at Berkeley. Mimeo.

Fershtman, C., Kamien, M. (1992) "Cross-Licensing of Complementary Technologies". International Journal of Industrial Organization 10(3), 329-348.

Galasso, A., Schankerman, M. (2008) "Patent Thickets and the Market for Innovation: Evidence from Settlement of Patent Disputes." Centre for Economics Policy Research Discussion Paper 6946.

Gallini, N. (1984) "Deterrence by Market Sharing: A Strategic Incentive for Licensing". American Economic Review 74(5), 931-941.

Gallini, N. (1992) "Patent Policy and Costly Imitation". The RAND Journal of Economics 23(1), 52-63.

Gallini, N. (2002) "The Economics of Patents: Lessons from Recent U.S. Patent Reform". Journal of Economic Perspectives 16(2), 131-154.

Gallini, N., Scotchmer, S. (2002) “Intellectual Property: When is it the Best Incentive System?” in A. Jaffe, J. Lerner and S. Stern, Eds., Innovation Policy and the Economy 2, 51-77.

Geradin, D., Layne-Farrar, A., Padilla, A.J. (2007) “The Ex Ante Auction Model for the Control of Market Power in Standard Setting Organisations”. Centre for Economic Policy Research Discussion Paper 6304.

Gilbert, R. (1995) “The 1995 Antitrust Guidelines for the Licensing of Intellectual Property: New Signposts for the Intersection of Intellectual Property and the Antitrust Law”. Department of Justice, Washington, D.C.

Gilbert, R., Katz, M. (2006) “Should Good Patents Come in Small Packages? A Welfare Analysis of Intellectual Property Bundling”. International Journal of Industrial Organization 24(5), 931-952.

Gilbert, R., Katz, M. (2007) “Efficient Division of Profits from Complementary Innovations”. Competition Policy Center Paper CPC07-076.

Gilbert, R., Shapiro, C. (1990) “Optimal Patent Length and Breadth”. The RAND Journal of Economics 21(1), 106-112.

Gilbert, R., Shapiro, C. (1997) “Antitrust Issues in the Licensing of Intellectual Property: The Nine No-No’s Meet the Nineties”. Brookings Papers: Microeconomics, 283-349.

Gilbert, R., Sunshine, S. (1995) “Incorporating Dynamic Efficiency Concerns in Merger Analysis: The Use of Innovation Markets”. Antitrust Law Journal 63, 569-602.

Gill, D. (2008) “Strategic Disclosure of Intermediate Research Results”. Journal of Economics and Management Strategy, *forthcoming.*.

Giuri, P., Mariani, M. Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., Garcia-Fontes, W., Geuna, A., Gonzales, R., Harhoff, D., Hoisl, K., Lebas, C., Luzzi, A., Magazzini, L., Nesta, L., Nomaler, O., Palomeras, N., Patel, P., Romanelli, M., Verspagen, B., (2006) “Everything You Always Wanted To Know About Inventor (But Never Asked): Evidence from the PATVAL-EU Survey”. Centre for Economic Policy Research Discussion Paper 5752.

Glachant, M., Meniere, Y. (2008) “Technology Diffusion with Learning Spillovers: Patent versus Free Access”. Ecole des Mines, Paris. Mimeo.

Green, J., Scotchmer, S. (1995) “On the Division of Profit in Sequential Innovation”. The RAND Journal of Economics 26(1), 20-33.

Guellec, D., van Pottelsberghe de la Potterie. (2007) The Economics of the European Patent System: IP Policy for Competition and Innovation. Oxford University Press: Oxford.

Hall, B. (2002) “The Financing of Research and Development”. Oxford Review of Economic Policy 18(1), 35-51.

Hall, B. (2005) “Exploring the Patent Explosion”. Journal of Technology Transfer 30(2_2), 35-48.

Hall, B. (2007) “Patents and Patent Policy” Oxford Review of Economic Policy 23(4), 541-567.

Hall, B., Graham, S., Harhoff, D., Mowery, D. (2003) “Prospects for Improving U.S. Patent Quality Via Postgrant opposition”. In A. Jaffe, J. Lerner and S. Stern, Eds., Innovation Policy and the Economy 4, 115-143.

Hall, B., Ziedonis, R. (2001) “The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995”. The RAND Journal of Economics 32(1), 101-128.

Harhoff, D., Henkel, J., von Hippel, E. (2003) “Profiting from Voluntary Information Spillovers: How Users Benefit by Freely Revealing their Innovations” Research Policy 32(10), 1753-1769.

Harhoff, D., Reitzig, M. (2004) “Determinants of Oppositions against EPO Patent Grants: The Case of Biotechnology and Pharmaceuticals”. International Journal of Industrial Organization 22(4), 443-480.

Harhoff, D., Wagner, S. (2005) “Modelling the Duration of Patent Examination at the European Patent Office”. Centre for Economic Policy Research Discussion Paper 5283.

Harhoff, D., Wagner, S. (2006) “Business Method Patents in Europe and their Strategic Use – Evidence from Franking Device Manufacturers” Discussion Paper in Business Administration 1265, University of Munich, Munich School of Management.

Heller, M., Eisenberg, R. (1998) “Can Patents Deter Innovation? The Anticommons in Biomedical Research”. Science 280(5364), 698-701.

Henkel, J. (2004) “The Jukebox Mode of Innovation – A Model of Commercial Open Source Development” Centre for Economic Policy Research Discussion Paper DP4507, July.

Hopenhayn, H., Mitchell, M. (2001) “Innovation Variety and Patent Breadth”. The RAND Journal of Economics 32(1), 152-166.

Hopenhayn, H., Llobet, G., Mitchell, M. (2006) “Rewarding Sequential Innovators: Prizes, Patents, and Buyouts”. Journal of Political Economy 114(6), 1041-1068.

Horowitz, A., Lai, E. (1996) “Patent Length and the Rate of Innovation”. International Economic Review 37(4), 785-801.

Horstmann, I, MacDonald, G., Slivinski, A. (1985) “Patents as Information Transfer Mechanisms: To Patent or (Maybe) Not to Patent”. The Journal of Political Economy 93(5), 837-858.

Hunt, R. (2004) “Patentability, Industry Structure, and Innovation.” Journal of Industrial Economics 52(3), 401-426.

Hylton, K. (2006) “Property Rules and Liability Rules, Once Again”. Review of Law and Economics 2(2), 137-191.

Jaffe, A. (2000) “The U.S. Patent System in Transition: Policy Innovation and the Innovation Process”. Research Policy 29, 531-557.

Jaffe, A., Lerner, J. (2006) Innovation and its Discontents: How our Broken Patent System is Endangering Innovation and Progress and What to Do About It. Princeton University Press: Princeton.

Janis, M. (2003) “Experimental Use and the Shape of Patent Rights for Plant Innovation”. University of Iowa School of Law. Mimeo.

Kaplow, L. (1984) “The Patent-Antitrust intersection: A Reappraisal”. Harvard Law Review 97(8), 1813-1892.

Katz, M., Shelanski, H. (2006) “Merger Policy and Innovation: Must Enforcement Change to Account for Technological Change?”. in A. Jaffe, J. Lerner and S. Stern, Eds. Innovation Policy and the Economy 1. 109-165

Katz, M., Shelanski, H. (2007) “Mergers and Innovation” Antitrust Law Journal 74, 1-86.

Katz, M., Shapiro, C. (1986) “Technology Adoption in the Presence of Network Externalities”. Journal of Political Economy 94(4), 822-841.

Katz, M., Shapiro, C. (1987) “R&D Rivalry with Licensing or Imitation”. American Economic Review 77(3), 402-420.

Kim, J., Marschke, G. (2005) “Labour Mobility of Scientists, Technological Diffusion, and the Firm’s Patenting Decision” The RAND Journal of Economics 36(2), 298-317.

Kitch, E. (1977) “The Nature and Function of the Patent System”. Journal of Law and Economics 20(2), 265-290.

Klemperer, P. (1990) “How Broad Should the Scope of Patent Protection Be?”. The RAND Journal of Economics 21(1), 113-130.

Kremer, M. (1998) “Patent Buyouts: A Mechanism for Encouraging Innovation”. Quarterly Journal of Economics 113(4), 1137-1167.

La Manna, M. (1994) "Research vs. Development: Optimal Patenting Policy in a Three-Stage Model". European Economic Review 38(7), 1423-1440.

Lakhani, K., Wolf, R. (2005) "Why Hackers Do What They Do? Understanding Motivation and Effort in Free/Open Source Software Projects". In J. Feller, B. Fitzgerald, S. Hissam and K. Lakhani, Eds. Perspectives on Free and Open Source Software. MIT Press: Cambridge, MA.

Langinier, C. (2005) "Using Patents to Mislead Rivals". Canadian Journal of Economics 38(2), 520-545.

Lanjouw, J., Lerner, J. (2001) "Tilting the Table? The Use of Preliminary Injunctions". Journal of Law and Economics 44(2), 573-603.

Lanjouw, J., Schankerman, M. (2001) "Characteristics of Patent Litigation: A Window on Competition". The Rand Journal of Economics 32(1), 129-151.

Lanjouw, J., Schankerman, M. (2004) "Protecting Intellectual Property Rights: Are Small Firms Handicapped?" Journal of Law and Economics 47(1), 45-74.

Layne-Farrar, A., Lerner, J. (2008) "To Join or Not to Join: Examining Patent Pool participation and Rent Sharing Rules". Harvard University, mimeo.

Lemley, M. (2001) "Rational Ignorance at the Patent Office". Northwestern University Law Review 95(4), 1497-1532.

Lemley, M., Shapiro, C. (2005) "Probabilistic Patents". Journal of Economic Perspectives 19(2), 75-98.

Lemley, M., Shapiro, C. (2007) "Patent Holdup and Royalty Stacking". Texas Law Review 85:1990-2049.

Lerner, J. (1995) "Patenting in the Shadow of Competitors". Journal of Law and Economics 38(2), 463-495.

Lerner, J. (2005) "150 Years of Patent Office Practice". American Law and Economics Review 7, 112-143.

Lerner, J., Tirole, J. (2002) "Some Simple Economics of Open Source". Journal of Industrial Economics 50(2), 197-234.

Lerner, J., Tirole, J. (2004) "Efficient Patent Pools". The American Economic Review, 94(3), 691-711.

Lerner, J., Tirole, J. (2005) "The Economics of Technology Sharing: Open Source and Beyond". The Journal of Economic Perspectives 19(2), 99-120.

Leveque, F., Meniere, Y. (2004) The Economics of Patents and Copyright. Monograph. Berkeley Electronic Press.

Lichtman, D. (2006) "Patent Holdouts in the Standard-Setting Process". Academic Advisory Council Bulletin 1.3. Progress and Freedom Foundation: Washington, D.C.

Manelli, A., Regibeau, P., Rockett, K. (1994) "Trademark Protection and Post-Patent Competition". Cuadernos Economicos de ICE 57(2), 115-139.

Matutes, C., Regibeau, P., Rockett, K. (1996) "Optimal Patent Design and the Diffusion of Innovations". The RAND Journal of Economics 27(1), 60-83.

Maurer, S., Scotchmer, S. (2002) "The Independent Invention Defence in Intellectual Property". Economica, New Series 69(276), 535-547.

Maurer, S., Scotchmer, S. (2004) "Procuring Knowledge". In Intellectual Property and Entrepreneurship, Advances in the Study of Entrepreneurship, Innovation and Economic Growth, Vol. 15, 1-31.

Maurer, S., Scotchmer, S. (2006a) "Profit Neutrality in Licensing: The Boundary Between Antitrust Law and Patent Law". American Law and Economics Review 8, 476-522.

Maurer, S., Scotchmer, S. (2006b) "Open Source Software: The New Intellectual Property Paradigm". In T. Hendershott, Ed. Handbook on Economics and Information Systems, Ch. 5. Elsevier: Amsterdam.

Menell, P., Scotchmer, S. (2007) "Intellectual Property" in M. Polinsky and S. Shavell, Eds. Handbook of Law and Economics. Elsevier: Amsterdam.

Merges, R. (1999) "Institutions for Intellectual Property Transactions: The Case of Patent Pools". University of California at Berkeley School of Law. Mimeo.

Merges, R.P., Nelson, R.R. (1990) "On the Complex Economics of Patent Scope". Columbia Law Review 90, 839-916.

Merges, R., Menell, P., Lemley, M. (2006) Intellectual Property in the New Technological Age, Fourth Edition. Aspen Publishers: New York.

Middendorf, W. (1981) What Every Engineer Needs To Know About Inventing. Volume 7 of What Every Engineer Should Know. CRC Press: Boca Raton, Fla.

Miller, A., Davis, M.H. (1983) Intellectual Property: Patents, Trademarks and Copyright in a Nutshell. West Publishing Co.: St. Paul, Minn.

Miyagiwa, K. (2007) "Collusion and Research Joint Ventures". The Institute of Social and Economic Research, Osaka University, Discussion Paper 704.

Nard, C., Morriss, A. (2006) "Constitutionalizing Patents: From Venice to Philadelphia". Review of Law and Economics 2(2), 223-321.

Noel, M., Schankerman, M. (2006) "Strategic Patenting and Software Innovation". Centre for Economic Policy Research Discussion Paper 5701.

Nordhaus, W. (1969) Invention, Growth and Welfare: A Theoretical Treatment of Technological Change. MIT Press: Cambridge, MA.

O'Donoghue, T (1998) "A Patentability Requirement for Sequential Innovation". The RAND Journal of Economics 29(4), 654-679.

O'Donoghue, T., Scotchmer, S., Thisse, J-F. (1998) "Patent Breadth, Patent Life, and the Pace of Technological Progress". Journal of Economics & Management Strategy 7(1), 1-32.

Pakes, A. (1986) "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks". Econometrica 54(4), 755-784.

Qian, Y. (2007) "Do Additional National Patent Laws Stimulate Domestic Innovation in a Global Patenting Environment? A Cross-Country Analysis of Pharmaceutical Patent Protection, 1978-2002". Review of Economics and Statistics 89(3), 436-453.

Regibeau, P., Rockett, K. (2007) "The Relationship Between Intellectual Property Law and Competition Law: An Economic Approach". In The Interface Between Intellectual Property Rights and Competition Policy, Steven Anderman, Ed., Cambridge University Press: Cambridge.

Sakakibara, M., Branstetter, L. (2001) "Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms". The RAND Journal of Economics. 32(1), 77-100.

Schankerman, M., Scotchmer, S. (2001) "Damages and Injunctions in the Protection of Intellectual Property". The RAND Journal of Economics 32(1), 199-220.

Scotchmer, S. (1991) "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law". The Journal of Economic Perspectives 5(1), 29-41.

Scotchmer, S. (1996) "Protecting Early Innovators: Should Second-Generation Products be Patentable?". The RAND Journal of Economics 27(2), 322-331.

Scotchmer, S. (1998) "R&D Joint Ventures and Other Cooperative Arrangements". In R. Anderson and N. Gallini, Eds. Competition Policy and Intellectual Property Rights in the Knowledge-Based Economy. Calgary: University of Calgary Press.

Scotchmer, S. (1999) "On the Optimality of the Patent Renewal System". The RAND Journal of Economics 30(2), 181-196.

Scotchmer, S. (2004). Innovation and Incentives. MIT Press: Cambridge, Ma.

Scotchmer, S., Gallini, N. (2002) "Intellectual Property – When is it the Best Incentive Mechanism?". In A. Jaffe, J. Lerner and S. Stern, Eds., Innovation Policy and the Economy 2, 51-78.

Scotchmer, S., Green, J. (1990) "Novelty and Disclosure in Patent Law." The RAND Journal of Economics 21(1), 131-146.

Segal, I., Whinston, M. (2007) "Antitrust in Innovative Industries". American Economic Review 97(5), 1703-1730.

Sena, V. (2004) "The Return of the Prince of Denmark: A Survey on Recent Developments in the Economics of Innovation". The Economic Journal 114, F312-F332.

Severinov, S. (2001) "On Information Sharing and Incentives in R&D". The RAND Journal of Economics 32(3), 542-564.

Shapiro, C. (2001) "Navigating the Patent Thicket: Cross-Licenses, Patent Pools and Standard-Setting". In A. Jaffe, J. Lerner, S. Stern, Eds. Innovation Policy and the Economy 1, 119-150.

Shapiro, C. (2003) "Antitrust Limits to Patent Settlements". The RAND Journal of Economics 34(2), 391-411.

Shapiro, C. (2004) "Patent System Reform: Economic Analysis and Critique". Berkeley Technology Law Journal 19(3), 1017-1047.

Shapiro, C. (2008) "Patent Reform: Aligning Reward and Contribution". In A. Jaffe, J. Lerner, S. Stern, Eds. Innovation and the Economy 8, 111-156.

Shavell, S., Ypersele, T. (2001) "Rewards versus Intellectual Property Rights". Journal of Law and Economics XLIV, 525-547.

Schmidt, K. (2008) "Complementary Patents and market Structure". Centre for Economic Policy Research Discussion Paper 7005.

Siebert, R., von Graevenitz, G. (2005) "How Licensing Resolves Hold-Up: Evidence from a Dynamic Panel Data Model with Unobserved Heterogeneity".

Sorenson, O., Fleming, L. (2004) "Science and the Diffusion of Knowledge". Research Policy 33(10), 1615-1634.

Swanson, D., Baumol, W. (2005) "Reasonable and Nondiscriminatory (RAND) Royalties, Standards Selection, and Control of Market Power". Antitrust Law Journal 73, 1-58.

Tao, Z, Wu, C. (1997) "On the Organization of Cooperative Research and Development: Theory and Evidence". International Journal of Industrial Organization 15(5), 573-596.

Tandon, P. (1982) "Optimal Patents with Compulsory Licensing". Journal of Political Economy 90(3), 470-486.

Trajtenberg, M. (2002) “Government Support for Commercial R&D: Lessons from the Israeli Experience”. In A. Jaffe, J. Lerner, S. Stern, Eds. Innovation Policy and the Economy 2, 79-134.

Van Dijk, T. (2000) “Licence Contracts, Future Exchange Clauses, and Technological Competition”. European Economic Review 44(8), 1431-1448.

Varian, H. (2005) “Copying and Copyright”. Journal of Economic Perspectives 19(2), 121-138.

Van Zeebroeck, N., Van Pottelsberghe de la Potterie, B. (2008) “Filing Strategies and Patent Value”. Centre for Economic Policy Research Discussion Paper 6821.

Von Graevenitz, G., Wagner, S., Harhoff, D. (2008) “Incidence and Growth of Patent Thickets – The Impact of Technological Opportunities and Complexity”. Ludwig Maximilians University, Munich School of Management. Mimeo.

Walsh, J., Arora, A., Cohen, W. (2003) “Effects of Research Tool Patents and Licensing on Biomedical Innovation”. In Cohen, W., Merrill, S., Eds. Patents in the Knowledge-Based Economy. Washington, D.C.: National Academies Press. 285-340.

Walsh, J., Cho, C., Cohen, W. (2005) “View from the Bench: Patents and Material Transfers”. Science 23 309(5743), 2002-2003.

Waterson, M. (1990) “The Economics of Product Patents”. The American Economic Review 80(4), 860-869.

Weston, R. (1998) “A Comparative Analysis of the Doctrine of Equivalents: Can European Approaches Solve an American Dilemma”. IDEA: The Journal of Law and Technology 39, 35-91.

Whinston, M. (1990) “Tying, Foreclosure, and Exclusion”. American Economic Review 80(4), 837-859.

Whinston, M. (2001) “Exclusivity and Tying in *U.S. v. Microsoft*: What We Know and Don’t Know”. Journal of Economic Perspectives 15 (Spring), 63-80.

Wright, B. (1983) “The Economics of Invention Incentives: Patents, Prizes, and Research Contracts.” The American Economic Review 73(4), 691-707.