Patent systems for encouraging innovation:
Lessons from economic analysis[1]

David ENCAOUA (CES, CNRS & University Paris I)[°]
Dominique GUELLEC (OECD)[x]
Catalina MARTÍNEZ (CSIC)[+]

Abstract

Economic theory views patents as policy instruments aimed at fostering innovation and diffusion. Three major implications are drawn regarding current policy debates. First, patents may not be the most effective means of protection for inventors to recover R&D investments when imitation is costly and first mover advantages are important. Second, patentability requirements, such as novelty or non-obviousness, should be sufficiently stringent to avoid the grant of patents for inventions with low social value that increase the social cost of the patent system. Third, the trade-off between the patent policy instruments of length and breadth could be used to provide sufficient incentives to develop inventions with high social value. Beyond these three implications, economic theory also pleads for a mechanism design approach: an optimal patent system could be based on a menu of different degrees of patent protection where stronger protection would involve higher fees, allowing self-selection by inventors.

Key words: patent, leading breadth, inventive step, sequential innovations, license, mechanism design.
JEL Classification: 031, 034

[°] Corresponding author, Université Paris I, MSE, 106-112 Boulevard de l'Hôpital, 75647 Paris Cedex, France tel: 00 33 1 44 07 81 97, fax: 00 33 1 44 07 82 31, e-mail: encaoua@univ-paris1.fr
[x] dominique.guellec@oecd.org, Organisation for Economic Cooperation and Development (OECD).
[+] catalina.martinez@csic.es, Consejo Superior de Investigaciones Científicas (CSIC), Instituto de Políticas y Bienes Públicos (IPP).

1. Introduction

There have been tremendous changes in the patent system over the past two decades, most of them going in the same direction: expanding and strengthening the protection of inventions[2]. The patent community, including attorneys, judges, patent officials and members of intellectual property business associations, has been a major driving force behind this evolution. Are these changes justified from an economic perspective based on what we have learned so far from research on the economics of patents? The answer is complex and depends on an assessment of whether the economic literature provides useful conclusions for what could be considered an optimal patent regime. Economists have only recently started to participate in policy discussions surrounding patents and to exchange views with the patent community in open fora; however there is still a long way to go and this paper aims at contributing to improving such communication[3].

Economic research in the area of patents is not new, but it has expanded and progressed considerably in recent years. The major justification given by practitioners for the existence and the working of the patent system refers to its effects on innovation and economic growth. However, economic analysis has not until very recently informed decisions taken by courts, patent offices and legislators. The traditional gap between economic research and patent policy might have been partially due to lack of communication between economic researchers and the patent community, whose members are mainly engineers and legal scholars, but it has also been due to the inability of economists to make operational their messages. Economic arguments are often cast in terms that are not especially helpful for policy makers, based on variables that do not constitute real policy levers. The purpose of this paper is to draw lessons from recent economic research on how the design of real policy levers in the field of patents may affect the innovation process.

We do not aim to present an exhaustive survey of recent research in the economics of patents and its relevance to all policy questions. We will instead focus on a limited number of key policy questions, leaving aside other important issues, such as what is the appropriate choice of patent regimes according to the state of development of a country, or the political economy of patent regimes.

---

[2] See e.g. Kortum and Lerner (1999), Jaffe (2000), Coriat and Orsi (2002), Gallini (2002), Martinez and Guellec (2004), Menell and Scotchmer (2005).

[3] Recent illustrations can be found in the 2002 FTC-DOJ Hearings on Competition and Intellectual Property Law and Policy in the Knowledge-Based Economy (http://www.ftc.gov/opp/intellect/index.htm) that led to the publication of a report by the FTC, 2003, as well as in the reports produced by the Committee on IPRs in the knowledge-based economy of the US National Research Council of the National Academies (NAS, 2003; 2004), the OECD conference on patents, innovation and economic performance held in Paris in August 2003 (OECD, 2004) and Guellec and van Pottelsberghe (2007) for an overview of the European scene.

Patent subject matter has expanded over the past decades to include biotechnology, software and, in some countries, methods of doing business. The extent to which patents are effective as an incentive mechanism in all fields of knowledge, that is, whether the incentives provided by competitive market mechanisms need to be supplemented or not by legal monopolies granted by governments to compensate inventors for their investment, is examined in the next section of the paper. As documented by empirical and theoretical studies, patents are necessary in certain, but not all cases. A note of caution should then be attached to the expansion of the domain covered by patents.

Some countries have arguably experienced a weakening of the standard criteria for granting patents (justified by the belief that "more patents is better"), and a tendency to grant patents with broader scope in certain technology fields (according to the principle that "broader patents are more valuable, then better") in the past few years. However, as reported in section 3, these trends might have detrimental effects on competition and follow-on invention and should be carefully monitored.

More fundamental reforms of the patent system as suggested by recent research are addressed in section 4 of this paper, such as the design of efficient patent protection mechanisms that provide strong incentives to invent while minimising the social costs associated with a monopoly position. This analysis is still highly theoretical and based on the idea that an efficient patent regime should encourage the self-selection of patentees for obtaining different degrees of patent protection, making the extension of their rights commensurable with the value of their invention to society.

Changes to patent regimes should be implemented carefully based on an analysis of their economic impacts on prices, on innovation and on diffusion. Emphasizing the use of patents as a policy instrument would reinforce their status by highlighting the benefits they bring to society and help prescribe a careful design and implementation of the rights they confer to inventors. However, the implementation of economic insights in the field of patent policy can be quite difficult in practice: it would require more empirical testing of the theoretical lessons reviewed here, together with improved communication and strong collaboration between economists and the patent community.

2. In what economic contexts are patents needed?

The cornerstone of the traditional economic argument in favour of patent protection is the non-rival character of knowledge, which means that once an invention is known, everyone can use it with no additional R&D cost. Economists have long challenged such an idea, notably from the evolutionary school (Nelson and Winter, 1982). They also have recently started to formalise arguments against this traditional view (Bester and Petrakis, 1998; Hellwig and Irmen, 2000; Boldrin and Levine, 2002; Quah, 2002). When imitating is as costly as inventing, or when firms have economic and technical means for protecting their inventions then, as the argument goes, there is no need for further legal protection. Under those circumstances, patents may simply become a source of market distortions and facilitate rent-seeking or strategic behaviour by patent holders.

2.1. Traditional arguments for and against patents

The usual argument in favour of intellectual property protection as it appears in the seminal works of Arrow (1962), Nordhaus (1969) and Romer (1990) is well known: innovation amounts to knowledge production, but knowledge is inherently non-rival (even when embedded in new products or technologies) which causes market failure and insufficient incentives to innovate.

The non-rival character of knowledge implies that the amount of knowledge available to any one user does not decrease when others use it. Knowledge consumption does not require any additional resources than those devoted to its initial production: once produced it can be subsequently used by others without its value being reduced. This non-rivalry property, satisfied by public goods, is to be contrasted with the rival character of private goods: my consumption of a private good reduces the total amount available for others. Knowledge has also a non-excludable character, in the sense that once it is produced, others cannot be stopped from benefiting from it, and as a result everyone can use it unless exclusive rights legally protect it.

In this context, it is traditionally argued, perfect competition in the product market does not allow innovators to recover their innovation costs when the production of knowledge requires the expense of a fixed and indivisible cost, in terms of R&D investment, and the goods and services in which the knowledge is embedded can be produced and distributed at low marginal cost. Public intervention is therefore needed to re-establish private incentives to engage in R&D activities and produce socially valuable knowledge, because its non-rival and non-excludable features cause market failure.

Patents have been generally considered a valid policy instrument to overcome such market failure, as an *ex-ante* incentive mechanism giving the inventor the exclusive right to use or sell its invention. By imposing a legal exclusivity on the use of knowledge, society encourages *ex-ante* investment in R&D and thus the production of knowledge and innovation. However, whereas for rival goods strong property rights lead to efficient market outcomes, for non-rival goods patents involve a trade-off. Weak rights may lead to under-provision of R&D, but strong rights may lead to an excessive monopoly distortion (deadweight loss) and to a slow-down in the pace of technical progress. Follow-on inventors may be confronted with obstacles raised by previous inventors, in terms of exclusive rights over knowledge they might need to access (Heller and Eisenberg, 1998). Patents appear to be a second best solution. The first best, characterised by a socially desirable level of innovation without market power and with global diffusion, appears to be unreachable.

Among the virtues of the patent system, the following properties must be emphasized. First, by giving some temporary exclusionary rights to inventors, the government delegates the R&D decision and leaves in the hands of the inventor the responsibility of recovering his R&D investment. Not only do individual agents have better information on the costs and benefits of R&D than the government, but delegating also

has the positive effect of avoiding moral hazard on the part of researchers, a problem that may be inherent in the implementation of other policy instruments such as ex-ante subsidies. Second, the assignment of costs is made to users rather than to tax payers. Third, in order to implement a patent system the government does not require sensible economic information that is only privately known, such as R&D cost and private value of the invention, thus avoiding adverse selection problems. The reward obtained by patent holders is linked to the private value of their patented inventions. Innovative firms weigh the cost of patenting against the value of their inventions when deciding whether to invest and patent. Finally, the information disclosure requirement of patents favours the diffusion of knowledge.

The following drawbacks or weaknesses of patents can also be mentioned, without the aim of being exhaustive. First, patents create static distortions corresponding to the classical deadweight loss that results from inefficient monopoly pricing: not all consumers valuing goods above their marginal cost can buy them. Second, the market reward from a patented good is not directly linked to the R&D cost needed to develop it. Moreover, inventors cannot fully capture the social value of their invention. Patents may provide insufficient incentives for inventors to develop socially valuable inventions because of positive spillovers of their ideas to other researchers. Third, patent races create some duplication of resources. Fourth, patents are much more oriented to creating substitutes than complementary goods and this leads to important problems of coordination unsolved by the patent system (Eswaran and Gallini, 1996). Fifth, patent enforcement requires a large amount of financial resources that are diverted from the innovation process itself.

2.2. Market-based means of protection for inventions

The view that market forces are not sufficient to compensate inventors has been challenged on different grounds for the past two decades in a series of empirical works. In contrast with the traditional consideration of knowledge as a public good, broad empirical evidence supports the view that knowledge used in economic processes is to a large extent difficult or costly to imitate. If knowledge was like any other ordinary rival asset, there would be no need for special legal protection and competitive rents would provide sufficient incentives to innovate. Firms would be able to charge their customers for the additional R&D cost incurred for the production of new knowledge embedded in their products and processes, without being undercut by lower cost imitators. The market power enjoyed by innovators in this context would thus be justified by the underlying market conditions and the nature of their specific technology, rather than by legal protection mechanisms.

At the heart of the debate opened by recent theoretical research is the question of whether market-based mechanisms that rely on standard competitive forces, with no special rights interfering, are sufficient for inventors to recover innovation costs. Boldrin and Levine (2002) define the *right of first sale* as the property right allowing an inventor to use his intellectual asset for either a productive purpose or for making profit from selling the first unit or prototype to a third party. This right of first sale is inherent in

all property rights. In contrast, exclusive rights going beyond the right of first sale and involving the right to control and limit the usage of the intellectual property are defined by Boldrin and Levine as *downstream protection* or *intellectual monopoly*. This second class of rights is the one conferred by patent protection.

Boldrin and Levine argue that the right of first sale may be sufficient to obtain revenues reflecting the full market value of the invention (*i.e.* the net discounted value of the future stream of consumption services generated by the first unit produced) and allow the inventor to recover his R&D investment. In their model, consumers are confronted with the following choice: at each date they can either consume the product or replicate it in order to sell a new copy. Consumers withdraw revenues from the inventor when they choose to replicate the product, and the inventor internalises this loss in his expected flow of revenues. By assumption, subsequent replication is characterised by non-increasing returns to scale (it takes time to replicate the original prototype and only a limited number of units can be replicated per period), so that the inventor may obtain sufficient infra-marginal or competitive rents to recover his R&D sunk cost. Therefore, in this setting, even without intellectual monopoly, inventors may have enough incentives to undertake R&D investments.

This result, however, depends strongly on certain critical assumptions of the model. First, for competitive rents to be sufficient to recover R&D costs, demand has to be elastic. When demand is elastic, if the cost of replication decreases over time due to the introduction of an improved technology for reproducing and diffusing the copies of the prototype, the increase in the number of prototypes sold by the inventor at each date would be larger than their price decrease and his revenue would thus increase. In this context, any reduction in the price increases profits, which makes the conclusion of the model straightforward. Second, the inventor is supposed to anticipate the residual demand for his product at each date, without any informational limitation, which is a rather stringent assumption. Third, the market value of the prototype would need to cover R&D costs, which might not be possible when initial R&D costs are too high, given the indivisibility of the initial R&D investment. Fourth, Quah (2002) shows that the results obtained by Boldrin and Levine do not hold in a continuous time framework, concluding that competitive markets are not sufficient to recover R&D costs in a 24/7 Internet-like scenario.[4]

Moreover, despite its apparent appeal against intellectual monopolies, the mechanism proposed by Boldrin and Levine might be equivalent to protecting inventions with secrecy, and be thus quite commonly used in the real world. When inventors do not diffuse their inventions and protect them by trade secrecy law, customers have the right to reverse engineer the product and replicate the invention, which indicates that this could be the advocated diffusion technology in the model developed by Boldrin and Levine. The question then is whether patents are preferable to secrecy. On the one hand patents favour the diffusion of

---

[4] Bester and Petrakis (1998) obtain similar results to Boldrin and Levine in a partial equilibrium framework, and Hellwig and Irmen (2000) in a general equilibrium framework.

knowledge due to their disclosure requirements, which means that the cost of replication may be lower for a patented invention than for an invention that is kept secret. On the other hand, secrecy prevents disclosure but allows the working of competitive markets. Anton and Yao (2004) show that, when property rights offer only limited protection, the value of the disclosure is offset by the increased threat of imitation. Since imitation depends on inferences the imitator makes about the innovator's advance, it appears that patenting is not always the best choice for firms. More precisely, large inventions are protected primarily through secrecy when property rights are weak.[5]

From an empirical perspective, industrial surveys show that first mover advantages, secrecy and the existence of complementary assets are generally preferred by firms to protect their inventions to legal means of protection like patents, although not to the same extent across sectors. The 1983 Yale Survey and the 1994 Carnegie Mellon Survey indicate that US manufacturing firms tend to use private appropriation mechanisms, such as exploitation of lead time and the use of complementary sales, manufacturing and service capabilities, in addition to secrecy and patents, to capture and protect the competitive advantage provided by innovations. Patents appear to be relatively more effective only in industries such as medical equipment and drugs, special purpose machinery, computers and auto parts, but are relatively less emphasised than other appropriability mechanisms in almost all industries, significantly below complementary sales and services (see Levin *et al*., 1987 and Cohen *et al*., 2000).

These surveys also provide evidence of patents yielding positive returns to patent holders. Using data from the 1994 Carnegie Mellon Survey, Arora *et al.* (2003) find that the additional payoff obtained from a patented invention relative to an unpatented invention (patent premium) differs largely across industries and is positive only in a few manufacturing industries, which are those where inventors patent most: drugs, biotech, medical instruments, machinery, computers and industrial chemicals. Mansfield (1986) obtains similar results based on a previous survey where US manufacturing firms were asked what fraction of inventions they would not have developed in the absence of patents between 1981 and 1983. He finds such fraction to be relatively high for pharmaceuticals (60%) and chemicals (40%), and very low for other sectors (less than 10% for firms in electrical equipment, primary metals, instruments, office equipment, motor vehicles and others).

As regards services, research undertaken to date indicates that patents may not be essential to promote innovation: the benefits provided by lead time advantage, complementary assets and network externalities might be sufficient for inventors to recoup R&D investment costs. Banks and other financial institutions have developed new products and processes over the past four centuries in the absence of legal protection

---

[5] See Encaoua and Lefouili (2006)

mechanisms. Financial methods only started to be extensively patented in the United States after the State Street Bank decision in 1998.[6]

An empirical study by Tufano (1989) on the incentives to innovate in financial markets shows that underwriting spreads on first offerings were not much larger than on late offerings for a sample of financial services introduced by investment banks between 1974 and 1986. The study also shows that although lead-time was relatively short for innovators, they usually enjoyed higher market shares than followers.[7] Consistent with these empirical findings and additional information from interviews with credit derivatives dealers and developers in investment banks, a theoretical model developed by Herrera and Schroth (2000) concludes that the incentive of investment banks to launch new financial products comes from the *informational advantage* provided by being the first in the market. Lead-time is important because it provides an informational advantage in a setting where learning by doing is crucial, rather than due to the short-term monopoly profits it may confer to the innovator before its competitors enter the market. In fact, Herrera and Schroth argue that the threat of imitation prevents the innovator from charging monopoly profits in the learning stage, when he is the only supplier. The innovator is able to recover investment costs because his informational advantage and expertise enables him to charge higher prices and enjoy larger market shares than his competitors at the competitive stage. Information is essential in a market where profit depends on risk management. Although the design of a financial product is fairly easy to imitate, its optimal exploitation can only be imperfectly imitated because it requires some expertise that is carefully protected by the innovator with secrecy.[8]

Recent theoretical research and broad empirical evidence thus challenge the traditional view about the failure of market-based mechanisms to protect inventions. Replicating an existing invention is costly and time consuming because knowledge is to a large extent embedded in individuals, in firms, and in physical goods and equipment. When knowledge is not codified, imitation is delayed and limited and gives rise to

---

[6] Since then USPTO has issued patents related to financial inventions on asset valuation, debt management, education finance, mortgages, privatisation, risk assessment, stock picking and working capital finance, among others (Thomas, 2001).

[7] Recent empirical findings by Carrow (1999) confirm that leaders tend to have larger market shares than followers but show that prices decrease as the number of competitors increases when the reputation of sellers and specific characteristics of the products compared are introduced into the analysis.

[8] One of the persons interviewed by Herrera and Schroth, a J.P. Morgan credit derivatives trader, explained that the economic value of a financial product lies in information about its performance, rather than in its design: *"Everybody can see the laid-out contract but what I am very careful not to disclose are the positions in my book. With this information you could track down the logic and see where I make money. Without it you could not price correctly the product, break down the risks involved, and understand what the components are. New ideas are not easily imitated: the developing process is a set of complex skills that are not easy to acquire".*

first mover advantages that could allow inventors to recover R&D costs. However, knowledge can be characterised as blueprint in scientific areas where discoveries are summarised in formulas, and in certain industries where knowledge is science-based (*e.g.* pharmaceuticals). Moreover, codifying know-how within firms is a recommended practice economy-wide to facilitate the internal exploitation of knowledge.[9]

In any case, lead-time in the market can be sustained if further learning guarantees the initial inventor permanent advantage over followers. Secrecy can also lengthen the duration of the first mover advantage and reduce the ability of competitors to improve upon initial inventions, and it is under certain conditions protected by trade secrecy law. The need to have access to complementary assets in order to obtain profits from a particular invention may also provide an advantage to the innovator with respect to imitators. Such complementary assets may include marketing (*e.g.* distribution networks, brand name, and reputation), manufacturing facilities and specific competencies. Firms not endowed with these assets face barriers to enter the market and compete on unequal terms with the innovator.

2.3. Patents and sequential innovation: complementarity

When innovation is sequential in the sense that an invention directly follows up on previous ones, the exclusive rights provided by patents may impede access to the knowledge embedded in previous inventions and slow down technological progress. Some authors argue that this is the case for software-related inventions, an area where patents are increasingly used by firms, in addition to copyright and market-based mechanisms which were until recently the only mechanisms traditionally used to recover investment costs in the software industry.

Bessen and Maskin (2002) argue that software firms are better off when they are imitated by competitors because imitation increases the probability of competitors achieving further innovations from which they can in turn benefit at later stages: "*When innovation is sequential and complementary, standard reasoning about patents and imitation may get turned on its head. Imitation becomes a spur to innovation, while strong patents become an impediment*".[10] They conclude that even if the initial rents earned by an innovator in the absence of patents may be lower than with patents, the benefits that accrue to him when he is allowed in his turn to build around the next innovation made by a competitor may outweigh the current loss.

---

[9] We thank an anonymous referee for pointing out that the characterisation of knowledge as a blueprint also has advantages for inventors as it facilitates the exploitation of knowledge within firms.

[10] *"By 'sequential' we mean that each successive invention builds on the preceding one – in the way that the Lotus 1-2-3 spreadsheet built on VisiCalc, and Microsoft's Excel built on Lotus.. And by 'complementary', we mean that each potential innovator takes a different research line and thereby enhances the overall probability that a particular goal is reached within a given time"* (Bessen and Maskin, 2002, p.4).

Their result strongly relies on the complementarity assumption according to which the probability of subsequent inventions is higher when more firms enter the market with new ideas, an assumption that is justified insofar as the existence of different lines of research increases the probability of discovery.

Two situations might arise in the absence of patent protection: either all firms invest in R&D or only one firm invests and the others imitate. In contrast, patent protection allows one of the firms to block entry to subsequent markets and delay subsequent innovations. The pace of innovation would be lower in the presence of patents, as an effect of complementarity, given that it is only in the absence of patent protection that all firms would be allowed to stay in the market and invest in R&D.

Bessen and Maskin conclude that patent protection would give rise to efficient R&D if and only if it is socially optimal to have a single firm investing in R&D, and assert that *generally* having more than one firm is socially efficient. They apply these results to the software industry, based on data showing a lack of increase in R&D intensity of top US software patentees after the expansion of patent subject matter to software in the United States in the 1980s, and conclude that patents are not a good instrument to spur software innovation insofar as they pre-empt firms from benefiting from other firms' inventions by creating monopolies and foreclosing market entry.

The model and its application to the software industry raise several problems however. First, they assume that patents create a perpetual monopoly, meaning that the patent holder can exclude entirely and forever competitors from the market. This is a rather extreme view of patents. Patents have limited breadth, which allows competitors to circumvent them, provided they incur the R&D costs eventually needed. Irreversibility of monopoly positions is a rather linear view of technical change, which excludes radical leaps by which the basic technology is changed and new players emerge. In addition, the technology disclosure role that patents are supposed to play should not be overlooked, as it aims at limiting or reversing their exclusionary effect. Second, their model does not recognise other means of protection, especially proprietary source code protected by copyright and trade-secrecy law and promoted by network effects, which have proved more exclusionary than patents in certain segments of the software market. If the matter is to promote competition, given that complementarity makes it socially beneficial to have several firms doing research, it remains to be proven that patents in general are worse than appropriation mechanisms based on secrecy. It might be the case in certain circumstances, but in other circumstances the reverse might hold: further research is needed in this area.

2.4. Patents and sequential innovation: essential inventions

The potential harm of patents may extend beyond a slow-down in the pace of innovation towards blocking whole new lines of research when second generation inventions cannot be achieved without accessing the knowledge embedded in first generation inventions. In this context, initial inventions may be considered as "essential facilities" in the sense that there is no possibility to invent around them to proceed with further research. Basic discoveries, genetic material and research tools in general may lie in this category.

A particular invention, lacking economic value on its own, may be essential as a research input for highly valuable applications in the future. Such inventions are labelled "research tools". Walsh, Arora and Cohen (2002) provide empirical evidence on the impact of patenting and licensing research tools in the field of biomedical innovation. The authors note that advances in molecular biology, automated sequencing techniques and bio-informatics have led biomedical research to increasingly depend on prior scientific discoveries or research tools. They also note that the role of universities has become more important in the past two decades, both as sources of inventions and as the origin of start-up firms, originally created as university spin-offs based on patents.

Despite the obstacles posed by patents in an area where scientific discoveries were until the 1980s traditionally placed in the public domain, they note that researchers seem to have developed 'working solutions' that *"combine taking licenses, inventing around patents, infringement (often informally invoking a research exemption), developing and using public tools and challenging patents in court)."* The question is whether the benefits of patenting outweigh the cost in terms of resources needed to find such working solutions.

2.5. Conclusion: patents should not be the solution by default

Patents are not needed as an incentive mechanism for all types of inventions. A series of characteristics of technologies and markets such as the gains from market-lead, the arrival rate of innovative ideas, the ease of imitation and the more or less marked cumulative character of innovation, determine both whether market-based means of protection provide sufficient incentives for innovation and the impact of patents on the pace of innovation. The optimal level of patent protection may differ across fields, with different solutions applying to industries as diverse as pharmaceuticals, software and finance.

However, patents serve other purposes different from excluding competitors from the market, something that should also be taken into account when evaluating the relevance of patent as a policy instrument. Patents are used as an argument in negotiations for cross-licensing agreements, as a signalling mechanism for shareholders, banks, venture capitalists, competitors or customers. They also contribute to social welfare by facilitating the diffusion of knowledge through information disclosure and by allowing the development of markets for technology. When an invention can be easily kept secret, as in the case of process innovations, a patent might be socially beneficial for the diffusion of knowledge even if it is not necessary as an incentive to innovate. These other private and social motives for patenting that may lead to social welfare gains seem to differ across technology fields and to a large extent economic theory falls short of addressing them.

3. Instruments of patent policy design

In what follows we consider some policy issues in the debate surrounding patents. Although they do not exhaust all the problems raised by the evolution of the patent system, they are nevertheless of prime importance. Recent trends in the patent system may indicate a weakening of patentability requirements

and an increase of patent breadth in certain jurisdictions and technology fields. The purpose of this section is to inquire whether these changes in patent policy are justified from an economic perspective.

3.1 Patentability requirement: inventive step

Patent offices grant patents to inventions complying with the patentability criteria of industrial application, novelty, and inventive step.[11] The inventive step is a technical concept meaning that the patented invention should not be obvious to a person skilled in the relevant art, for instance amounting to a combination of existing techniques.

Expert estimates of the grant rate at USPTO indicate that it reached levels of around 90% during the 1990s (Quillen and Webster, 2001). Some observers have also argued that a large number of patents have been granted in the United States to inventions that were neither non-obvious nor novel, especially in new areas such as software. The increase of the workload at patent offices caused by a boom in patent applications, mainly in new areas such as biotechnology and software, has gone hand in hand with growing pressures to reduce the time lag between application and grant, and might have led to less careful examination standards. Even though most of this evidence remains anecdotal and partial, a general sentiment of a relaxation of patentability requirements prevails in certain jurisdictions (FTC, 2003; NAS, 2003 and 2004; OECD, 2004)

Economic research can provide some insight on the impact of lowering patentability requirements, although it might at first appear difficult to apply its lessons in practical terms. Not only the approach taken by economists towards patentability requirements differs largely from the practice at patent offices, but the latter might also differ in its daily implementation from the standards reflected in statutes and legal precedent.

Patent examiners define the inventive step according to the technical character of the invention, by reference to a person with ordinary skills in the art, whereas economists interpret it as a one-dimensional scale on which all techniques of the relevant domain can be measured, according to some performance feature (*e.g.* cost, quality). Economists define the inventive step by the threshold below which the reduction in costs for process innovations or the degree of quality improvement for product innovations would be insufficient for the patent to be granted. One difference between these two approaches is that the technical characteristics of an invention are usually known *ex-ante*, whereas its economic features are only known *ex-post*, after the technique has been actually implemented and has given rise to products sold on the market. A new technical device which does not improve performance with respect to a currently used technique may nevertheless satisfy the patentability requirement on the examiners' scale, even if it is not

---

[11] The US patentability criterion of ´utility´ would correspond to the European criterion of 'industrial application', whereas the US ´non-obviousness´ criterion would correspond to the European criterion of 'inventive step'.

acceptable *prima facie* on the economists' scale because it provides no direct or immediate cost reduction or quality improvement.

Traditional tests of non-obviousness at patent offices are primarily based on the following technical features: i) scope and contents of prior art; ii) differences between prior art and patent claims; iii) level of ordinary skills in the relevant art. However, in the United States, a set of secondary economic factors have been recently emphasised in some decisions of the US Court of Appeals for the Federal Circuit, such as commercial success, failure of others and long felt need (Hunt, 1999). EPO also considers commercial success as relevant information, but subordinate to technical criteria: "*Commercial success alone is not to be regarded as indicative of inventive step, but evidence of immediate commercial success when coupled with evidence of a long-felt want is of relevance provided that the examiner is satisfied that the success derived from the technical features of the invention and not from other influences (e.g. selling techniques or advertising).*"[12] Merges (1992) notes that the use of commercial success as a test of non-obviousness may have the drawback of rewarding firms with better marketing and distribution systems rather than to those introducing more innovative products in the market.

What is the advantage of imposing non-obviousness requirements? The answer is that strong patentability requirements provide higher incentives to innovate, either by extending the effective life of patents, that is, the length of market incumbency for the inventor, or by increasing the quality of the successive innovations (O'Donoghue, 1998; Hunt, 1999; O'Donoghue and Zweimuller, 2000).

To analyse the effect of the patentability requirement on the pace of innovation, Hunt (1999) considers a situation where successive quality improvements of an exogenous size occur randomly. At each point in time, only the highest quality product is profitable ('winner takes all') and an improvement is only patented if it exceeds some threshold value, which corresponds to the patentability requirement. In addition, an invention that does not fulfil this requirement falls in the public domain so that the previous patent holder can immediately copy it (secrecy is not an option in this model). A patented technology is fully disclosed and serves as a basis for further improvements. A fixed number of firms race to improve existing products subject to decreasing returns on R&D.

In this setting, the net effect on the pace of innovation of raising the patentability threshold depends on the magnitude of two opposing effects. The short-term effect is negative: increasing the patentability requirement lowers the probability that an invention qualifies for a patent, reducing therefore the short-term incentive to innovate. In contrast, the long-term effect is positive for two reasons. First, increasing the patentability requirement leads to longer effective patent lives for inventions reaching the threshold, by delaying its substitution in the market by an improved technique. Second, it leads to a higher average profit flow from a patented discovery. Hunt (1999) shows that the balance of these two opposing effects is

---

[12] EPO guidelines, Part C, Chapter IV, page 72.

in favour of the positive effect, insofar as the patentability threshold is not too large. He shows that there exists an inverted U relationship between the patentability requirement and the rate of innovation. The rate of innovation is first positively affected by the height of the patentability requirement but beyond a certain level the relationship changes and the rate decreases. Hunt (2003) obtains the same result by considering that industry structure is endogenous, so that the number of firms investing in R&D depends on the fixed cost of setting up an R&D lab.

Either with endogenous or exogenous industry structure, Hunt finds that not only is there a unique patentability standard that maximizes the rate of innovation in an industry, but also that it is stricter for industries where the arrival rate of ideas is high. In contrast, in an industry with a low rate of arrival of ideas, the negative short-term effect (fewer patentable inventions) of a marginal increase in the non-obviousness requirement is likely to be larger than the positive long-term effect (higher average value of patentable inventions), which may lead to less innovative activity or fewer firms entering the patent race in a model with free entry, since less firms will be able to amortize the costs of setting up an R&D lab.

O'Donoghue (1998) obtains a similar result when the size of innovations is endogenous. In this case, not only does a stringent patentability requirement extend the effective life of patents but it also induces firms to pursue more ambitious R&D projects. He also shows that the optimal patentability requirement increases with the arrival rate of innovative ideas and argues that a higher standard for protection does not lead to a decline of the rate of innovation in rapidly innovating industries. Moreover, increasing the patentability requirement in these industries may encourage larger and riskier inventions, something that is socially desirable when markets tend to favour smaller inventions with more certain rewards. Merges (1992) also argues that patents should be used to encourage firms to engage projects with low certainty of commercial success, as inventions with more certain gains would be implemented, even in the absence of patent protection. Conversely, reductions in the patentability requirement are more likely to encourage innovation in industries that innovate slowly.

Optimal patentability requirements are higher when technical change is more rapid or innovative ideas arrive more frequently because in that case the length of the incumbency period is shorter, and thus the opportunity cost of not getting a patent decreases. An implication of this finding is that the factors affecting the optimal level of the patentability requirement are technology specific, whereas the current patent system is characterised by uniform rules, according to the 'one size fits all' principle. The reduction in the patentability requirement experienced in certain countries over the past two decades might have favoured industries where technical change is slow at the expense of more dynamic areas. This is all the more important since patentability requirements seem to be rather low in new areas of patentability such

as biotechnological or software-related inventions, generally characterised by high frequency of arrival of innovative ideas where the optimal patentability requirements should be higher.[13]

Moreover, if imitation were assumed to be costly even for a fully disclosed invention, the positive effect on innovation of increasing patentability requirements would be reinforced and the optimal level of the patentability requirement would be higher. Indeed, if the incumbent cannot freely capture a non-patented technology, he would need more incentives to undertake an R&D project. Increasing the patentability requirement would give such an incentive by extending the effective life of the incumbent's patent.

When innovation is cumulative, the issue of ensuring that each innovator is sufficiently rewarded, taking into account the benefits conferred to future innovators, is of a different nature. As noted by Scotchmer, "quality ladders present a different incentive problem than the problem of basic and applied research" (Scotchmer, 2004, p.157). When innovation is cumulative and ideas occur randomly, there is no distinction between first-generation and second-generation innovators. It has been argued that choosing an appropriate patent breadth is a better policy lever than determining a required inventive step (Scotchmer, 2004). However, when one uses another creative environment framework in which ideas are costly (i.e. R&D spending creates knowledge outputs), it appears that the determination of a threshold for the inventive step encourages firms "to be more ambitious than they otherwise would be" (Scotchmer, 2004, p.150).

3.2. Patent breadth and the effective life of patents

A patent holder has the right to prevent others from making, selling or using the invention protected by the patent. In principle, patent breadth is determined by the claims accorded by the patent examiners to the patentee, defining the boundaries between what is protected and what is not, and by the courts' interpretation of these claims where litigation occurs. In addition, courts may rely on doctrines that either expand patent breadth, such as the *doctrine of equivalents*, or restrict it, such as the *enablement* doctrine. Whereas the *doctrine of equivalents* states that essentially equivalent products not explicitly included in the patent claims could be judged to infringe the patent, the *enablement doctrine* states that the patent protects only what is disclosed. The existence of discrepancies between the technical scope of inventions and the claims granted by patent protection is a currently debated issue, notably in biotechnology (Bar-Shalom and Cook-Deegan, 2002) and software.

In the real world, whereas patent length has a uniform statutory content across industries and countries, patent laws do not include any formal concept of breadth. The fine-tuning of patent design that can be achieved in economic models is therefore more difficult to reach in reality. Nevertheless, there is a

---

[13] The rate of technical change may itself be a function of patentability standards, which complicates the story. We thank and anonymous referee for pointing this out.

presumption that, under the constraint on the uniform statutory patent life, broad patents may do more bad than good to innovation, as broad patents may distort incentives and allocation of research funds.

The profitability of a patent is a function of its breadth (which determines the flow of revenue in each period of time) and its duration (how long the flow lasts). While broad patents increase the rents accruing to inventors, they might also generate rent-seeking behaviours if the patent office is not entirely successful in screening the legitimate claims of each patent application. Broad patents increase the social cost of imperfections in the management of the patent system. In addition, broad patents tend to skew the reward distribution associated with research. Large breadth makes research resemble more a "winner takes all" game, as modelled in patent races. One drawback of such a scheme is that it may lead to duplication and concentration of R&D efforts in some areas at the expense of investment in other areas where the return is lower.

Another conclusion, emerging from the literature examined below, is that the trade-off between patent breadth and length depends on industrial characteristics such as the level of technological opportunities or the rate of arrival of innovative ideas. Both patent breadth and length determine the extent of patent protection and affect the expected revenues from patenting an invention, but they work in different ways, with different effects on the economic behaviour of the patent holder and his competitors. Whereas broad patents allow the patent holder to set a higher market price for the patented product, patents with longer lives allow patent holders to obtain revenues for longer periods of time. Larger breadth makes it more difficult to imitate or improve upon the protected invention, whereas increasing the duration of patent protection enhances the incentives to imitate or to improve the invention (Gallini, 1992).

The effects of patent breadth and length differ according to whether an innovation is considered in isolation or as part of a sequence of innovations building on each other. The following two sections summarise economic research related to these two settings.

3.2.1 Isolated innovations

Seminal works on patent design emphasize the fundamental trade-off between dynamic efficiency and static welfare loss (Nordhaus, 1969; Scherer, 1972). The resolution of this trade-off led to the economic justification for a finite length of protection, according to the argument that it is better to forego some innovations by restricting patent life in order to reduce deadweight loss on those that are realised. The trade-off between patent length and patent breadth has been investigated in more recent works. Both longer protection or larger breadth stimulate R&D but exacerbate the deadweight loss. Optimal patent breadth is obtained by minimising the discounted value of the deadweight loss created by the patent under the constraint that the discounted profit provides enough incentives to invest.

It appears that the optimal design depends heavily on both the retained definition of breadth and the way patent breadth affects deadweight loss (Tandon, 1982; Gilbert and Shapiro, 1990; Klemperer, 1990; Waterson, 1990; Gallini, 1992; Denicolò, 1996). In the isolated innovation framework, patent breadth is

identified with the concept of *lagging breadth*, understood as the protection accorded against imitators. For a product innovation, lagging breadth defines the minimal differentiation between the patented product and an inferior product for the latter not to infringe the patent of the former. For a process innovation, lagging breadth defines the maximal cost reduction allowed to a non-infringing inferior process.

Gilbert and Shapiro (1990) define patent breadth as the minimum price that provides the patentee a pre-specified reward allowing him to recover the R&D cost (participation constraint). In this case, minimising the deadweight loss under the participation constraint calls for a long and narrow patent where the patent breadth just allows the specified reward. This result crucially depends on the fact that the deadweight loss function increases more than proportionally with breadth (i.e. it is a convex function of breadth). Tandon (1982) obtains the same qualitative result in the context of compulsory licensing.

Klemperer (1990) obtains more mitigated results by developing a model with horizontal product differentiation, where patent breadth is defined as the set of characteristics in the product space that infringe the patented product. Consumers are deprived of their ideal products if these products happen to infringe on an existing patented product. In this context, Klemperer shows that either a configuration involving long and narrow patents or a configuration with short and broad patents can be optimal.

Short and broad patents are also found to be optimal in Gallini (1992) by using a different notion of breadth. In her model, the breadth of a patent determines how costly it is to develop a non-infringing substitute for the protected product. Since a larger breadth is associated with a higher cost, breadth should be set at a level that deters wasteful duplication of research (finding a close substitute). For patents of a given breadth, duration has to be adjusted accordingly. In this setting, the social cost of a patent has two components: the deadweight loss during the patent life and the cost incurred by the development of a substitute. Gallini shows that short and broad patents are optimal because they avoid socially wasteful costs to develop substitutes. However, this result has been criticised by Maurer and Scotchmer (2002) on the grounds that licensing could be an alternative way to avoid the development cost of such substitute products. When the cost for developing a substitute is close to the R&D cost of the original innovator, then high enough for deterring wasteful duplication, the previous result is reversed when licensing is allowed: long and narrow patents are optimal. However, even when licensing is allowed, broad and short patents remain optimal when the cost of developing a substitute is sufficiently low, as a broader patent protecting the original invention increases the level of investment needed to develop a substitute without infringing, hence deterring such development by competitors.

To conclude, even if no definite results stem from this literature on the optimal mix between length and breadth for isolated innovations, there is a strong presumption that a combination of narrow and long protection is preferable, when other features such as patent races, licensing and characteristics of the product market competition are introduced (Denicolò, 1996).

3.2.2. Sequential innovations

The aim in a sequential innovation setting is to design a patent system that increases the rate of innovation. Suppose that ideas for improving a current technology occur randomly and the level of improvement is itself random. Not only must the initial and subsequent innovators receive sufficient rewards to recover their costs but also these rewards must take into account the fact that each innovation opens the way to further improvements. By deciding whether an improvement infringes or not a previous patent, the courts determine the effect of patents on the rate of innovation.

O'Donnoghue *et al.* (1998) show that in a cumulative setting, lagging breadth alone (even when combined with an infinite statutory patent length) does not provide sufficient incentives for R&D. Lagging breadth offers protection only against imitators and not against future innovations; therefore, no future innovation will infringe a previous patent in a regime where forward protection is absent. As a consequence, the concept of *leading breadth* is introduced in the setting, and it represents the extent of forward protection that limits the behaviour of subsequent innovators.[14]

They also analyse the problem of the optimal mix of leading patent breadth and patent length for sequential innovations. A specified rate of innovation can be achieved by one of the following alternative patent regimes: either a long statutory patent life combined with a narrow leading breadth or a short statutory patent life combined with a broad leading breadth. Two different views lead to a justification of either one of these two regimes. Under the first regime, based on the view that the basic invention might have been achieved by serendipity or as a result of a patent race, effective patent life is determined by the time elapsed until a sufficiently better product is invented and is thus dependent on the extent of the leading breadth. Under the second regime, effective patent life coincides with the statutory length of the patent and no new patentable invention is allowed before the expiration date of the first patent. Any improvement of the basic patented technology is an infringement and the leading breadth simply acts as an inventive step requirement. The second patent regime would be therefore equivalent to a situation with a high patentability requirement and short patent length. However, patentability requirements and leading breadth affect the pace of innovation in different ways, and so would these two patent regimes. Increasing the inventive step (second patent regime) may impede the appearance of a useful improvement of the basic invention, while increasing the leading breadth (first patent regime) has not necessarily such a negative impact, as long as the patent holder of the basic invention accepts to enter into a licensing agreement (Denicolò and Zanchettin, 2002).

Beyond these interpretative views, the comparison between the two patent regimes proposed by O'Donoghue *et al.* (1998) depends on the rate of arrival of innovative ideas (supposed to be exogenous). In industries where the rate of arrival of innovative ideas is low, the first patent regime (long statutory life

---

[14] Note that leading breadth elicits investment in infringing inventions, insofar as they can be licensed to the market incumbent.

with modest breadth) would be superior because it reduces market distortions and lowers R&D costs. Conversely, in industries where the rate of innovation is high, the second patent regime (large leading breadth and short length) would be more effective since it gives higher incentives for R&D. Moreover, a high rate of innovation increases the turnover of innovators. To sum up, one practical conclusion that can be retained from O'Donnoghue *et al* (1998) is that, as in the isolated case, the cumulative setting exhibits a trade-off between the patent's length and the extent of its claims (patent breadth). Unfortunately, such a trade-off does not seem to be considered in patent statutes or court enforcement at present.

3.3 Access to patented knowledge by subsequent inventors

When there are two successive innovations, the key is to find some distribution of profits between both innovators so that, not only each of them receives sufficient rewards to recover her own costs, but also the first innovator captures a part of the value created by the second innovation. When the initial innovation gives rise to a subsequent innovation, either as an application or as an improvement, one source of revenues for the first inventor is the license fee that the second inventor must pay, where her invention infringes. But, *ex-post*, the amount of the fee fixed by the first innovator may be sufficiently high to put the subsequent innovator in jeopardy, where the improved product infringes. Two alternative solutions may be used to solve this difficulty. The first consists of letting both innovators enter into *ex-ante* contractual agreements to share their expected joint profits. The second is to introduce a compulsory licensing regime, in which the first innovator commits to sell a license at a predetermined price, before the subsequent innovation is achieved.

Ex-ante contracting, before incurring the R&D costs needed to develop the improvement, can reproduce the situation where a single firm develops both initial and subsequent innovations (Scotchmer and Green, 1990; Merges and Nelson, 1990, 1994; Green and Scotchmer, 1995; Scotchmer, 1991, 1996, 2004). The introduction of *ex-ante* contracting is a simple application of the Coase theorem, which states that in the absence of transaction costs, the distribution of property rights among market participants has no effect on the social outcome. The advantage of *ex-ante* contracting is that subsequent innovations occur only if they provide additional profit, and the role of leading breadth is to determine how the profit is divided. However, this solution is difficult to implement, first because the second innovator is generally unknown to the first one ex-ante, and second because an ex-ante contractual agreement may be considered as a violation of competition rules.

Chang (1995), Denicolò (2002), and Denicolò and Zanchettin (2002) address these issues. According to Chang (1995), a simple cut-off rule, often used by the Courts, where a dividing line separates small improvements that infringe and large improvements that do not, may be a misleading solution. Measuring the first patent's breadth as the share of the private value of the second innovation appropriated by the first innovator, Chang (1995) argues that the optimal patent breadth is non-monotonic on the value of the first innovation. "*Courts should set infringement standards so as to extend the broadest protection not only to*

a basic invention with a very large stand-alone value relative to all possible subsequent improvements, but also to the patent with very little stand-alone value relative to the improvements that it may inspire." Chang provides also very limited support in favour of collusive licensing, showing that it creates incentives for inefficient entry by imitators who invent around the original patent. He advocates that given the impossibility of reaching a first best solution, a second best solution could be achieved under a compulsory licensing regime that regulates the price of the license. Otherwise, with the limited set of actual policy instruments available, only a third best level could be achieved. Without compulsory licensing, courts can only decide whether the second invention infringes or not the first patent, and subsequently, whether to allow collusive licensing or not.

Denicolò (2002) and Denicolò and Zanchettin (2002) suggest a more lenient policy against collusion between successive patent holders (for instance through cross-licensing agreements). In a model involving patent races for two successive innovations in which the winner of the first race can participate in the second one, they show that when social value differs from private value, collusion may be socially beneficial in some circumstances. In particular, if the second-generation innovation contributes to consumers' surplus more than the first, collusion between successive innovators should be permitted since it encourages the first invention.

It appears that licensing strategies of patent holders are as relevant as granting policies by patent offices for the impact of patents on innovation in the cumulative setting. Not only should patent offices avoid granting excessively broad patents on basic inventions that may be used on a wide range of applications in the future, but also competition authorities should monitor the potential anticompetitive effects of licensing agreements. The antitrust treatment of restrictive conditions included in licensing contracts is crucial. For instance, the US Antitrust Guidelines for the Licensing of Intellectual Property issued in 1995 make clear that the restraints in licensing contracts must be challenged under the rule of reason, so that authorities "*will consider whether the restraint is reasonably necessary to achieve pro-competitive efficiencies*". They also recognise that cross licensing or patent-pooling arrangements provide "*pro-competitive benefits by integrating complementary technologies, reducing transaction costs, clearing blocking positions and avoiding costly infringement costs*" (OECD, 1998).

4. Rethinking the patent system: patents as an incentive mechanism

The question of patent policy design is addressed in economic research from an *ex-ante* perspective, with the view to induce private agents to devote sufficient efforts and resources to undertake innovations that increase social welfare. In this setting, some attempts have been made to deepen the understanding of the relative virtues and shortcomings of different incentive systems to stimulate innovation (Merges and Nelson, 1990, 1994; FTC, 2003; Sideri and Giannoti, 2003; Scotchmer, 2004).

The existence of asymmetric information between firms and public authorities on the cost and value of research programs justifies the existence of patent protection. If public authorities were as well informed

as firms, *ex-ante* public sponsorship or *ex-post* prizes would be better incentive mechanisms than patents. Market distortions would be avoided and public authorities would be able to ensure the implementation of projects generating large social benefits (Wright, 1983; Shavell and Van Ypersele, 2001; Chiesa and Denicolò, 2002; Scotchmer, 2004).

Economists have attempted to understand the conditions under which patents can be considered as an efficient mechanism and, by doing that, they have also sought to design patents that elicit socially beneficial inventions while minimising distortions in the allocation of resources. Two major results arise from this recent literature.

For an isolated innovation, the patent system in which a longer patent life is chosen by the patentee in return for a higher fee may be an efficient system when neither the R&D cost nor the private value of the invention is observed by the patent office. *Renewal fees* can be used to implement an optimal mechanism of that kind (Scotchmer, 1999; Cornelli and Schankerman, 1999).

For successive innovations, the patent system may be also an efficient incentive mechanism when two inter-related policy instruments are introduced: a *fee on patent breadth* and a *buyout price* at which the patent holder commits to sell his rights to follow-on inventors building on his invention (Llobet *et al*., 2000). This research shows that, when designed in a way that allows *self-selection* by inventors, the patent system may be an incentive-compatible mechanism. However, research in this area is still at an early stage and results rely on strong assumptions regarding technology and informational structures.

4.1. Renewal fees as an incentive-compatible mechanism for isolated innovations

In many countries, including Europe and United States, it is necessary to pay renewal fees in order to keep the patent in force year after year until a maximum of 20 years from filing. Although in practice renewal fees appear to serve the purpose of covering administrative costs at patent offices, some economists have used information on patent renewals to make inferences about the distribution of the value of patented inventions (Lanjouw, 1998; Pakes, 1986; Lanjouw, Pakes and Putnam, 1996; Schankerman, 1998) and others have explored the use of renewal fees as policy instruments to design patents as incentive-compatible mechanisms (Cornelli and Schankerman, 1999; Scotchmer, 1999).

In general, information on the characteristics of a particular innovative project is more readily available to the firm than to the patent office, namely the R&D costs and the expected private value per period (i.e. the expected monopoly profit per period of time that the inventor can capture). From the firm's perspective, investment in a particular innovative project is profitable if and only if R&D costs are lower than some threshold varying positively with the private value of the innovation. This threshold defines a *cut-off* function giving the maximum R&D cost that can be supported by the firm in order to undertake a project leading to an expected private value per period. In turn, the public authority needs to design a direct revelation mechanism that implements the firm's decision rule efficiently, so that when the firm is asked to report its characteristics, it is induced to report the true parameters (incentive compatible mechanism).

In this framework, Scotchmer (1999) defines a patent-type mechanism for an isolated innovation characterised by two parameters: the price to be paid to obtain protection (lump-sum patent fee) and the corresponding length of protection acquired (patent-length). This mechanism is described as a menu of different pairs of patent fees and lengths, where fees are a non-decreasing function of length. Each patent applicant chooses a pair from this menu according to his R&D cost and the expected monopoly profit per period. Thus, applicants face a trade-off between either paying a higher patent-fee to obtain a longer patent life and capturing a higher overall profit or paying a smaller patent fee (or even receiving a subsidy from the public authority in some cases) for a shorter patent life protection. Applicants with high value innovations may want to pay for longer protection, even if benefiting from a longer protection is made more costly.

The lump-sum patent fee is set in such a way that whenever R&D costs are equal to the cut-off value, the patent mechanism enables the innovator to break-even. In contrast, when R&D costs are lower than the cut-off value, the mechanism lets the firm capture an extra-profit corresponding to the *informational rent*, and if R&D costs are larger than the cut-off value, the project is not started. With a convex cut-off function, both the patent length (equal to the slope of the cut-off function) and the lump-sum fee are increasing functions of the reported value. As a result, higher-value inventions will benefit from a longer patent life because inventors will pay higher fees.

From the social welfare point of view, this scheme is only justified if R&D costs increase more than proportionally with the value of the invention or, in other words, if R&D cost is an increasing and convex function of the value of the innovation. This is exactly the case where the patent system is an incentive-compatible mechanism. Thus, Scotchmer obtains two important results: First, only decision rules where the maximum R&D cost increases more than proportionally with the value of the project can be implemented by an incentive-compatible mechanism. Second, this incentive-compatible mechanism is equivalent to a patent mechanism.

These results give a theoretical justification to the patent system by presenting an economic environment in which the patent system turns out to be optimal and opens the way for rethinking how patent fees should be calibrated with the objective of increasing economic efficiency. From a normative perspective, the objective of public authorities would be to design a socially efficient decision rule, considering that patented inventions affect social welfare positively through the monopoly profits perceived by the inventor and the consumer surplus raised by the commercialisation of the invention, and negatively through the deadweight loss derived from monopoly pricing in the product market. Thus, the optimal cut-off function to be implemented by the patent system gives the maximum R&D cost that sustains an efficient project from a social welfare perspective.

Scotchmer (1999) also shows that any incentive compatible mechanism implementing an efficient R&D investment decision rule by using a lump-sum patent fee and the length of patent protection as policy

instruments can be replicated by a mechanism using non-decreasing *renewal fees* and the length of protection as alternative policy instruments. A scheme of patent renewal fees is an important instrument for an optimal patent policy. Increasing renewal fees may act as a 'sorting device', so that only patent holders with the most valuable innovations would have incentives to pay higher fees in order to extend the effective life of their patents. A system of renewal fees corresponds to an option-based view of patent valuation in the sense that, as long as the patent is renewed, the firm reveals the expected valuation of its discovery (Pakes, 1986).

Cornelli and Schankerman (1999) obtain similar results by using a simpler framework than Scotchmer (1999).[15] They compute the welfare gains that would be obtained from transforming an optimal uniform patent policy into an optimally differentiated patent policy, and compare situations in France, Germany and the United Kingdom.[16] Concerning welfare gains, simulations indicate an average welfare gain of 5%, which may exceed 10% for some R&D productivity parameters. As regards optimal patent lives, the optimal uniform patent length is between 15 and 19 years, close to the 20 years of statutory patent length currently in place in most countries, but optimally differentiated patent lengths for a range of R&D productivity parameters appear to be much more contrasted. First, there is a minimum length of about 7 years, even for very low R&D productivity. Second, for projects with very high productivity involving the greatest contribution to social welfare, the optimal length may be much longer than existing statutory length. Third, the range of optimal patent lives is between 8 and 15 years for the bulk of the R&D productivity distribution.

Finally, simulations indicate that optimal renewal patent fees must rise sharply with patent life and more rapidly than the associated profits from the patent. According to the authors *"this important feature of the optimal mechanism is violated by existing renewal schedules"*. Using the estimates on the value of patent rights obtained by Schankerman and Pakes (1986), Cornelli and Schankerman (1999) show that the equivalent tax rates implied by current patent fees in the three countries decline sharply with the effective patent life, leading to a *regressive tax*.[17] In turn, the tax rates that derive from the optimal patent mechanism would increase with the chosen patent life, leading to a *progressive tax*.

---

[15] The model of Scotchmer corresponds to an adverse selection problem involving two unknown parameters, the cost and the value, whereas the model of Cornelli and Schankerman combine adverse selection and moral hazard dimensions, each one with a single unknown parameter. However, the moral hazard variable is related to a productivity parameter that depends on the unobserved cost. So, the model of Cornelli and Schankerman amounts to an adverse selection model involving only one unknown parameter.

[16] Renewal fees are currently required for patent ages of 2-20, 3-18 and 5-16 in France, Germany and the UK respectively.

[17] The equivalent tax rate is defined by the ratio between actual cumulative renewal fees and the profits from patent protection.

4.2. The buyout mechanism for sequential innovations

The problem of designing an optimal incentive-compatible mechanism is much more complex when innovation is considered as a cumulative process, as exclusive rights on a particular invention might be used to prevent other inventors to improve upon it (see Section 3.2.2).

Llobet, Hopenhayn and Mitchell (2000) have made an important contribution in this area. They characterise the optimal mechanism as a patent with no statutory expiration date that provides a certain 'amount of protection' against future improvements. As in Scotchmer (1999) for the case of isolated innovations, the mechanism proposed by Llobet *et al*. for cumulative innovations offers a higher patent protection in exchange of a higher fee, the difference being that in the cumulative innovation context protection is defined in terms of *breadth* rather than in terms of *length*. Paying a higher fee for longer protection is not sufficient to design an incentive-compatible mechanism when innovation is cumulative.

The main contribution made by Llobet *et al* (2000) is to show that an incentive-compatible mechanism can be implemented in a very practical way, with only two instruments. Namely, a *mandatory buyout price* defining the minimum price at which a patent holder commits to sell his rights to any new innovator improving upon his innovation and accepting to pay such a price, and a *patent fee* that increases with the buyout price. By offering a menu of positively correlated *buyout prices* and *patent breadth fees*, public authorities are able to generate information on the innovations. An inventor with a more valuable invention will seek a broader patent, even at the cost of a higher fee, since reporting an underestimated value would expose him to having his invention purchased by a third party at a lower buyout price. Thus, the buyout price determines the breadth of the patent. The higher the buyout price the larger the implied breadth of the patent, because future innovators would need substantial improvements to be willing to pay a higher buyout price.

The main virtues of this mechanism are twofold. First, it delegates the choice of patent breadth to the applicants and it does so in an efficient way. Second, "*it generates enough information to solve the infringement problems through self-selection from a menu of patents. This is in contrast to current policy, where the courts must determine more than just if two innovations are quality improvements over the other, but also how much of an improvement has been made*" (Llobet *et al*, 2000, p.4). In the proposed mechanism where innovators choose the desired leading breadth by paying an appropriate fee and committing to a buyout price, the level of improvement at which infringement occurs is determined by the innovators themselves, lowering the burden on patent offices and courts.

Once more it appears that, even in the more complex context of sequential innovations, the introduction of a mechanism that gives the innovators sufficient incentives to reveal their private information could lead to a substantial improvement relative to the 'one size fits all' principle that currently characterises the patent system.

In reality, the patent system offers more diversity and flexibility than the "uniformity principle" advocated by legal scholars would suggest, leaving to patent applicants and holders some room for self-selection. One example is the "second tier patent protection", which may take the form of *utility models*, *petty patents* or the so-called *innovation patents* recently introduced in Australia. These are titles granted to inventors without examination: they are cheap, but offer less protection than real patents (shorter period of protection, more fragile in courts) and are explicitly targeted at small inventions. A second example is the European patent system itself, with its two tiers: the national level, with national patent offices granting patents of national validity, and the EPO granting Europe-wide patents. The latter is more expensive, but has wider market coverage and usually requires a more stringent inventive step (implying larger breadth) than the former. A third example is the PCT procedure: a type of application which allows the applicant to benefit from provisional protection while postponing up to 18 months his decision to file an actual patent application or not (a decision which triggers higher expenses). These different mechanisms have not been intended to serve society, but instead they are aimed at increasing the comfort of applicants (petty patents and PCT), or simply result from history (the European system). Still, they play a role of self-selection, and could be used as starting points by policy authorities for re-designing a patent system that would have better economic properties.

5. Conclusions and policy recommendations

This overview of the economics of patents and patent policy has underlined a series of practical issues that deserve further attention from policy makers.

Patents are a double-edged sword, with a positive and a negative side. They often contribute to enhancing incentives to invent, to disclosing and trading technology, but they also generate costs to society in terms of monopoly rents and barriers to access and use of knowledge.

Competitive rents, in the absence of patent protection, might be sufficient to compensate innovators in certain circumstances. For instance, when secrecy is a feasible means of protection and the cost of imitation if high, or first mover advantages and network externalities are important, patents may not be necessary to encourage innovation. However patents could still play a positive role for fostering disclosure (as opposed to secrecy) and market transactions over technology (licenses). Industry specific conditions prevail in that matter.

Patents with low social value, or even illegitimate patents (not novel or not sufficiently inventive), can have a detrimental effect on innovation and competition. Patentability requirements should thus be kept high and be strictly applied. Special attention should be given to new technology fields, notably software and biotechnology, where there is not yet an established examination tradition in patent offices, to allow monitoring of granting standards.

Patent fees should reflect the cost of patents to society, rather than patent offices' examination costs. Application and renewal fees could be used as self-selection mechanisms to encourage high valuable

inventions to be patented and discourage the least valuable ones. The funding structure of patent offices should thus be changed. Government should see patent offices not as profit centres, but as agencies in charge of aspects of innovation policy.

Economic research has taught us that a good patent system should be flexible and make extensive use of the price mechanism for patentees to self-select their desired protection according to the characteristics of their inventions. How this can be implemented remains to be carefully investigated, but should at least be looked into by policy makers and institutions whose aim is to improve the effectiveness of patents as an innovation policy tool.

REFERENCES

Aghion P., Harris, C. and J. Vickers (1997), "Competition and growth with step-by-step innovation: an example", *European Economic Review*, 41, 771-782.

Anton, J. and D. Yao (2004), "Little patents and big secrets: managing intellectual property", *Rand Journal of Economics*, 35, 1, 1-22

Arora, A., M. Ceccagnoli and W.C. Cohen (2003), "R&D and the patent premium", NBER Working Paper n.9431.

Arrow, K.J. (1962), "Economic welfare and the allocation of resources for invention" in Universities-National Bureau of Economic Research Conference. Series, The rate and direction of economic activities: economic and social factors. Princeton University Press.

Bar-Shalom, A. and R. Cook-Deegan (2002), "Patents and innovation in cancer therapeutics: lessons from CellPro", *The Milbank Quarterly*, Vol. 80, No. 4.

Bessen, J. and E. Maskin (2002), "Sequential innovation, patents and imitation", MIT Working Paper Department of Economics, revised version.

Bessen, J. and R. Hunt (2003), "An empirical look at software patents", http://www.researchoninnovation.org/swpat.pdf

Bester, H. and E. Petrakis (1998), "Wage and productivity growth in a competitive industry", CEPR Discussion Paper 2031.

Boldrin, M. and D.K. Levine (2002), "The case against intellectual property", forthcoming in *Papers & Proceedings* of the *American Economic Review*, May 2002.

Carrow, K.A. (1999), "Evidence of early-mover advantages in underwriting spreads", *Journal of Financial Services Research*, 15, 1, 37-55.

Chang, H. (1995), "Patent scope, antitrust policy and cumulative innovation", *The Rand Journal of Economics*, 26, 1, 34-57

Chiesa, G. and V. Denicolò (2001), "Patents, prizes and optimal innovation policy", W.P. University of Bologna

Cohen, W. M., Nelson, R.R. and J. P. Walsh (2000), "Protecting their intellectual assets: appropriability conditions and why US manufacturing firms patent or not", NBER Working Paper 7552.

Coriat, B. and Orsi F. (2002), "Establishing a new Intellectual Property Rights Regime in the United States: Origins, Content and Problems", *Research Policy*, Vol. 31, n°7-8.

Cornelli F. and M. Schankerman (1999), "Patent rewards and R&D incentives", *The Rand Journal of Economics*, 30, 2, 197-213

Denicolò V. (1996), "Patent races and optimal patent breadth and length", *The Journal of Industrial Economics,* XLIV, 3, 249-265

Denicolò V. (2000), "Two-stage patent races and patent policy", *The Rand Journal of Economics*, 31,3, 488-501.

Denicolò V. (2002), "Sequential innovation and the patent-antitrust conflict", Mimeo, Department of Economics, University of Bologna

Denicolò V. and P. Zanchettin (2002), "How should forward protection be provided"?, *International Journal of Industrial Organization,* 20, 801-827

Encaoua D. and A. Hollander (2002), "Competition policy and innovation", *Oxford Review of Economic Policy*, 18, 1, 63-79

Encaoua D. and D. Ulph (2000), "Catching-up or leapfrogging: the effects of competition on innovation and growth", *Cahiers de la MSE, EUREQua*, 00017, revised version 2004

Encaoua, D. and Y. Lefouili (2006), "Choosing intellectual protection: Imitation, Patent Strength and Licensing", CESifo WP 1715, forthcoming in *Annales d'Economie et de Statistique*

Eswaran M. and N. Gallini (1996), "Patent policy and the direction of technical change", *The Rand Journal of Economics*, 27,4, 722-746

European Commission (2002), "Proposal for a Directive of the European Parliament and of the Council on the patentability of computer implemented inventions", Brussels, 20.02.2002, COM(2002) 92 final.

FTC (2003), To Promote innovation: the proper balance of competition and patent law and policy, A report by the Federal Trade Commission, October 2003, http://www.ftc.gov/os/2003/10/innovationrpt.pdf

Gallini, N. (1992),"Patent policy and costly innovation", *The Rand Journal of Economics*, 21, 106-112.

Gallini, N. and S. Scotchmer (2002), "Intellectual property: when is it the best incentive system?", in Innovation Policy and the Economy, eds. A. Jaffe, J. Lerner and S. Stern, MIT Press.

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

Gans, J., Hsu, D.H. and S. Stern (2002), "When does start-up innovation spur the gale of creative destruction?" *The Rand Journal of Economics*, Vol. 33, No.4.

Gilbert, R. and C. Shapiro (1990), "Optimal patent length and breadth", *The Rand Journal of Economics*, 21, 106-112.

Green, J. and S. Scotchmer (1995), "On the division of profit between sequential innovators", *The Rand Journal of Economics,* 26, 20-33.

Guellec D. and B. van Pottelsberghe (2007), "The Economics of the European Patent System". Oxford University Press, Oxford.

Hall, B. and R.H. Ziedonis (2001), "The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995", *The Rand Journal of Economics*, Vol.32, No.1, pp.101-128.

Harhoff, D., Regibeau, P. and K. Rockett (2001), "Genetically modified food. Evaluating the economic risks", *Economic Policy*, October 2001, CEPR.

Heller, M.A. and Eisenberg, R.S. (1998), "Can patents deter innovation? The anticommons in biomedical research", Science 280, 698-701.

Hellwig, M. and A. Irmen (2000), "Endogenous technological change in a competitive economy", *Journal of Economic Theory*, 101, 1-39.

Herrera, H. and E. Schroth (2000), "Profitable innovation without patent protection: the case of credit derivatives", Working Paper at the Berkeley Center, Stern NYU.

Hunt R.M. (1999), "Nonobviousness and the Incentive to Innovate: An Economic Analysis of Intellectual Property Reform",Federal Reserve Bank of Philadelphia, Working Paper 99-3.

Hunt, R.M. (2003), "Patentability, industry structure, and innovation", Federal Reserve Bank of Philadelphia, Working Paper 01-13/R.

Jaffe, A. (2000) "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process," Research Policy 29(4-5), April 2000, pp. 531-557.

Klemperer, P. (1990), "How broad should the scope of patent protection be?", *The Rand Journal of Economics*, 21, 113-130.

Kortum S. and J. Lerner (1999), "What is behind the recent surge in patenting?" Research Policy 28(1) January 1999 pp. 1-22.

Lanjouw (1998), "Patent protection in the shadow of infringement simulation estimations of patent value", Review of Economic Studies, 65, 4, 671-710

Levin, R.C., Klevorick, A.K., Nelson, R.R. and Winter, S.G. (1987), "Appropriating the returns from industrial R&D", Brookings Papers on Economic Activity: 783-820.

Llobet G., H. Hopenhayn and M. Mitchell. (2000), "Rewarding sequential innovators, patents, prizes and buyouts", Federal Reserve Bank of Minneapolis, Research Department Staff Report, 273

Mansfield, E. (1986), "Patents and innovation: an empirical study", *Management Science*, 32, 173-181.

Martinez C. and Guellec D. (2004), "Overview of recent changes and comparison of patent regimes in the United States, Japan and Europe", Chapter 7 in *Patents, Innovation and Economic Performance. OECD Conference Proceedings*, OECD, Paris.

Matutes, C., Rockett, K.E. and P. Regibeau (1996), "Optimal patent protection and the diffusion of innovation", *The Rand Journal of Economics*, 27, 60-83.

Maurer, S. M. and S. Scotchmer (2002), "The Independent Invention Defense in Intellectual Property", *Economica*, 69, 535-547.

Mazzoleni, R. and R.R. Nelson (1998), "The benefits and costs of strong patent protection: a contribution to the current debate", *Research Policy*, 27, 273-284.

Menell, P. and S. Scotchmer (2005), "Intellectual Property", forthcoming in *Handbook of Law and Economics*

Merges R.P. (1992), "Uncertainty and the Standard of Patentability", *High Technology Law Journal*, 7, 1-70.

Merges, R.P. and R.R. Nelson (1990), "On the complex economics of patent scope", *Columbia Law Review*, Vol.90, pp.839-916.

Merges, R.P. and R.R. Nelson (1994), "On limiting or encouraging rivalry in technical progress: the effect of patent scope decisions", *Journal of Economic Behaviour and Organisation*, 25, 1-24.

NAS (2003), *Patents in the knowledge-based economy*, Wesley M. Cohen and Stephen A. Merrill, Editors, Committee on IPRs in the Knowledge-Based Economy, National Research Council of the National Academies, Washington, D.C. http://www.nap.edu/catalog/10770.html

NAS (2004), *A patent system for the 21st century*, Stephen A. Merrill, Richard C. Levin, and Mark B. Myers, Editors, Committee on IPRs in the Knowledge-Based Economy, National Research Council of the National Academies, Washington, D.C. http://www.nap.edu/html/patentsystem

Nelson R.R. and S.G. Winter (1982), "An evolutionary theory of economic change", Cambridge, Harvard University Press

Nordhaus, W. (1969), *Invention, growth and welfare*, Cambridge, MA, MIT Press.

OECD (1998), *Competition policy and intellectual property rights*, DAFFE/CLP(98)18, http://www.oecd.org/FR/documents/0,,FR-documents-0-nodirectorate-PL-22-no-0,00.html

OECD (2004), *Patents, innovation and economic performance. OECD conference proceedings*, OECD, Paris.

O'Donoghue, T. (1998), "A patentability requirement for sequential innovation", *The Rand Journal of Economics*, 29, 654-679.

O'Donoghue, T., S. Scotchmer and J. Thisse (1998), "Patent breadth, patent life and the pace of technological progress", Journal of Economics and Management Strategy, 7, 1, 1-32

O'Donoghue T. and J. Zweimuller (2000), "Patents in a model of endogenous growth", CEPR D.P. 1951

Pakes, A. (1986), "Patents as options: some estimates of the value of holding European patent stocks", *Econometrica*, 54, 755-784.

Quah, D. (2002), "24/7 competitive innovation", W.P. *LSE*

Quillen, C.D. and O.H. Webster (2001), "Continuing Patent Applications and Performance of the U.S. Patent and Trademark Office", 11 Federal Circuit Bar Journal, 1, 1-21

Romer, P. (1990), "Endogenous technical change", Journal of Political Economy, 94, 5, 71-102

Schankerman, M. (1998), "How valuable is patent protection? Estimates by technology field", Rand Journal of Economics, 29, pp: 77-107.

Schankerman, M. and A. Pakes (1986), "Estimates of the value of patent rights in European countries during the post-1950 period", *The Economic Journal*, 96, 1052-1076.

Scherer, M. (1972), "Nordhaus theory of optimal patent life: a geometrical reinterpretation", American economic Review, 62, 3, 422-427

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

Scotchmer, S. (1991), "Standing on the shoulders of giants: cumulative research and the patent law", Journal of Economic Perspectives, Symposium on Intellectual Property Law, 5, 1, 29-41

Scotchmer, S. (1996), "Protecting early innovators: should second generation products be patentable?", *The Rand Journal of Economics*, 27, 322-331.

Scotchmer, S. (1999), "On the optimality of the patent renewal system", Rand Journal of Economics, 30, 2, 181-196

Scotchmer, S. and J. Green (1990), "Novelty and disclosure in patent law", *The Rand Journal of Economics*, 21, 131-146

Shapiro C. (2001), "Navigating the patent thicket: cross licenses, patent pools and standard settings", in A. Jaffe, J. Lerner and S. Stern (eds), *Innovation Policy and the Economy*, Cambridge, MA, MIT Press

Shavell, S. and T. Van Ypersele (1999), "Rewards versus Intellectual Property Rights", *Journal of Law and Economics*, XLIV, 525-547

Sideri S. and P. Giannoti (2003), "Patent system, globalization, and knowledge economy", CESPRI, Universita Commerciale Luigi Bocconi, W.P. 136

Tamai, K. (2001), "The experimental use exception: a Japanese perspective", International Perspectives on the Experimental Use Exception, CASRIP Symposium Publication Series, available at http://www.law.washington.edu/Casrip/

Tandon, P. (1982), "Optimal patents with compulsory licensing", *Journal of Political Economy*, 90, 470-486.

Thomas, J.R. (2001), "Patenting pricing on the nines? An overview of patents on financial services and other methods of doing business", mimeo, George Washington University Law School. http://www.phil.frb.org/econ/conf/innovations/thomas.pdf

Tufano, P. (1989), "Financial innovation and first mover advantages," *Journal of Financial Economics* 25 (1989), 213-240. Condensed version published as "Financial Innovation and First Mover Advantages," *Journal of Applied Corporate Finance* 5 (1992), pp. 83-87

Tufano, P. (2002), "Financial innovation", forthcoming chapter in The Handbook of the Economics of Finance (North Holland). http://www.people.hbs.edu/ptufano/

Van Dijk, T. (1996), "Patent height and competition in product improvements", *Journal of Industrial Economics*, 44, 151-167.

Walsh, J.P., Arora, A. and W.M. Cohen (2002), "The patenting and licensing of research tools and biomedical innovation", Science, Technology and Economic Policy Board of the National Academies, forthcoming.

Waterson, M. (1990), "The economics of product patents", *The American Economic Review*, 80, 860-869

Wright, B. (1983), "The Economics of invention incentives: patents, prizes, and research contracts", *The American Economic Review,* Volume 73, Issue 4, 691-707.