The Economics of Patents: An Overview

Corinne Langinier and GianCarlo Moschini

Working Paper 02-WP 293
February 2002

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070
www.card.iastate.edu

Corinne Langinier is adjunct assistant professor of economics and GianCarlo Moschini is professor of economics and Pioneer Hi-bred International Chair in Science and Technology Policy, both in the Department of Economics, Iowa State University.

This paper was prepared for inclusion in the volume *Intellectual Property Rights and Patenting in Animal Breeding and Genetics*, edited by Scott Newman and Max Rothschild.

For questions or comments about the contents of this paper, please contact GianCarlo Moschini, 583 Heady Hall, Iowa State University, Ames, IA 50011-1070; Ph: 515-294-5761; Fax: 515-294-6336; e-mail: *moschini@iastate.edu*.

Abstract

In this paper, we review the economic effects of intellectual property rights and specifically address the economics of the patent system. The production and dissemination of new knowledge is fraught with market failures because knowledge is a public good. Patents provide a second-best solution to the resulting appropriability problem. We review the main benefits and costs of the patent system, focusing on the role that patents play in providing incentives for innovation, in promoting the dissemination of knowledge, and in helping technology transfer and commercialization of new technology. From a more normative perspective, we address the questions of what the features of an optimal patent system are and whether the patent system is socially desirable. We examine the problem of the optimal length and scope of patent protection, both for the case of a single innovation and for the richer case of cumulative innovations. Finally, we review the issues related to how the patent system influences the market structure and research and development investments.

Key words: innovation, intellectual property rights, monopoly, patents, research and development.

THE ECONOMICS OF PATENTS: AN OVERVIEW

Introduction

Patents are perhaps the most important legal instruments for protecting intellectual property rights. A patent confers to an inventor the sole right to exclude others from economically exploiting the innovation for a limited time (20 years from the date of filing). To be patentable, an innovation must be *novel* in the sense of not constituting part of the prior art or more generally of not being already in the public domain. A patentable innovation also must involve an inventive step, meaning that it must be *non-obvious* to a person with ordinary skills in the particular field of application. The innovation also must be *useful* to be patentable; that is, it must permit the solution of a particular problem in at least one application. A major element of a patent application is *disclosure*: the invention must be described in sufficient detail to *enable* those skilled in the particular field to practice it. The patent application also lays out specific claims as to the scope of the patent itself. The traditional statutory scope of patents—encompassing machines, industrial processes, composition of matter, and articles of manufacture—excluded important kinds of scientific discoveries such as laws of nature, natural phenomena, and abstract ideas. But recent developments in the use of patents for computer software, information technology, and biotechnology innovations are challenging a reductive interpretation of such exclusions. See Merges (1997) for further details and a comprehensive treatment of U.S. patent law.

Given that private property is inherently associated with the freedom of choice that constitutes the cornerstone of market economies (Barzel 1989), it may seem obvious that intangible assets associated with human inventiveness and creativity should enjoy a legal status similar to that afforded to the ownership of other more standard goods and services. However appealing, this consideration does not quite do justice to the many aspects of the economics of patents. There are at least two relevant approaches one can take to articulate a fuller discussion of this intellectual property institution: a positive analysis and a normative one. From a positive perspective, we may ask how patents, as currently implemented, affect the workings of the economic system. Ultimately we wish to understand how the existence of patents affects the allocation of resources to, and the distribution of income arising from, inventive activities. A positive economic framework also can be used to ask the related question of why patents came to exist. Alternatively, from a more normative perspective, we can explore whether the institution of patents is a desirable feature of the economic system, and what the features of an optimal patent system are. In what follows, we consider these economic questions in more detail, focusing on conceptual and theoretical analyses.[1]

Benefits and Costs of Patents

From an economic perspective, the crucial features of patents are that (1) they deal with new knowledge, as embodied in an innovative product or process, and (2) they confer (limited) monopoly rights to the inventor. New knowledge that makes possible the production of new products and/or processes obviously carries considerable economic value, but it has features that make it problematic for the market system to handle properly (Arrow 1962). Specifically, knowledge is a quintessential public good. Pure public goods have two basic attributes. First, they are non-rival in consumption, meaning that a person's use of a public good does not affect the amount of it that is available for others. Second, they are non-excludable, meaning that it is not possible to prevent individuals from enjoying the public good once it is available. An example of a pure public good is national defense. It is clear that, absent intellectual property rights, most discoveries and inventions would exhibit public good attributes.

The problems that a competitive system has with public goods are readily apparent. An inventor may bear all the cost of an innovation, but everyone benefits (possibly to varying degrees) from a discovery, and thus everyone has an incentive to free ride on the innovative efforts of others. The inherent externalities associated with this class of public goods generate a market failure: a competitive market system may be expected to provide an inefficiently low level of innovations. Intellectual property rights in general, and patents in particular, address this problem by attacking the non-appropriability of knowledge that lies at the heart of this market failure. Specifically, by endowing innovators with property rights on their discoveries, patents are a legal means of affecting the excludability attributes of an otherwise pure public good.

The main economic benefits and costs of the patent system are intimately related to the nature of the market failure that it addresses, and to the second-best character of the solution that it provides. We begin the discussion of such effects along the taxonomy suggested by Mazzoleni and Nelson (1998).

Patents Can Promote New Discoveries

By endowing discoverers with property rights over the fruits of their efforts, patents affect the incentive to innovate and are likely to increase the flow of innovations. This increase is presumably desirable, given that otherwise the market system may provide too little new knowledge. But by giving the patentee exclusive rights on the exploitation of a unique economic good that is still non-rival in consumption, a patent creates a monopoly situation that adversely affects the efficient use of new knowledge.

Nordhaus (1969) and Scherer (1972) provided an early formalization of this fundamental trade-off between benefits and costs of the patent system. For a simpler illustration, consider the potential market for a new product, say, a new pharmaceutical. If p denotes the price charged to consumers, the latent demand $D(p)$ is likely to display a choke price $\overline{p}$ such that $D(p)=0$ for all $p \geq \overline{p}$ (there exists a prohibitive price that drives demand to zero). Furthermore, this latent demand is downward sloping, i.e., $D'(p)<0$ for all $0 \leq p < \overline{p}$ (a lower price makes the drug attractive to an increasing number of potential users). Suppose that it takes a total cost F to develop this drug, including the testing required for approval, but once this knowledge is available, it takes only a unit cost c to actually manufacture and sell the drug. This situation is represented in Figure 1, where demand applies for a specific period (a year, say). Given some qualifications, area $(S+\Pi+L)$ represents the per-period “social surplus,” i.e., the monetary benefits that would accrue to society if this new product were developed (and produced at the efficient level q^C). If the present discounted value of the current and future stream of such benefits exceeds the research and development (R&D) cost F, then net benefits are possible for society from the introduction of the new drug. But clearly, absent intellectual property rights, no individual has an incentive to incur the cost F: the ability of competi-

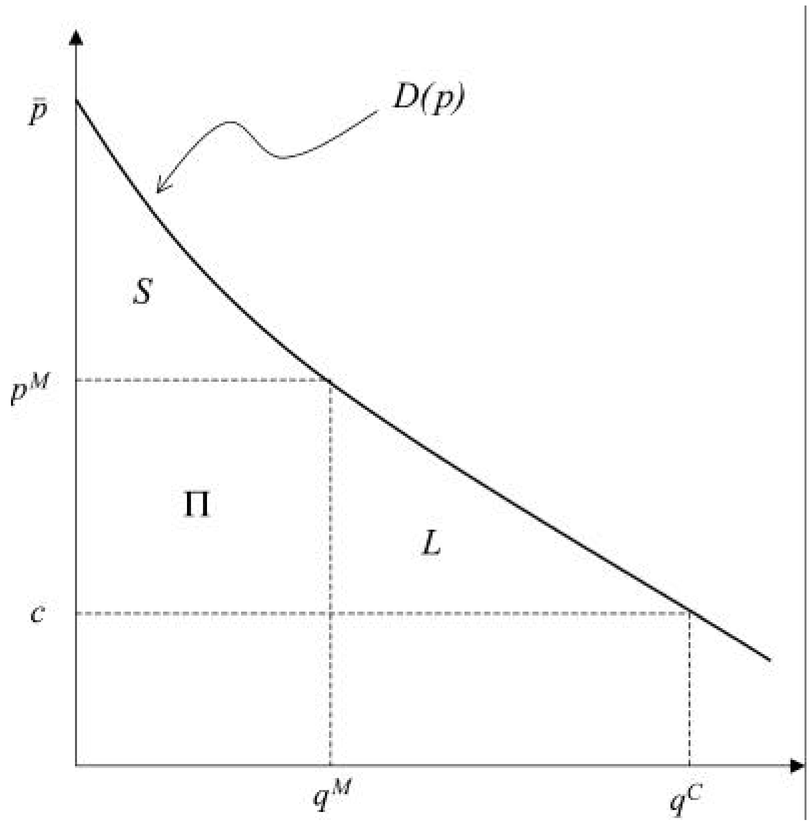

FIGURE 1. Patents and the market for a new product

tors to readily copy the innovation would drive the market price to its unit production costs c, and the cost F cannot be recouped.

But with a patent, the innovator becomes a monopolist in the market for the new drug and can profit by pricing the drug optimally at p^M, where $c < p^M < \bar{p}$. The innovator's per-period profits are represented by the area Π in Figure 1. If the present discounted value of such profits, over the life of the patent, exceeds the R&D cost F, then a sufficient incentive exists for this innovation to be brought about, and society as a whole enjoys the benefits $(S + \Pi)$ for the duration of the patent, and benefits $(S + \Pi + L)$ thereafter. But note that, during the life of the patent, the innovation is produced at an inefficiently low level from a social point of view (i.e., $q^M < q^C$).[2] This brings to the fore a fundamental trade-off of the patent system: the balancing between the benefits of encouraging additional innovative activities and the costs of forgoing the competitive provision of some goods and services. *Ex post*, that is, given that an innovation is available, a monopoly is bad from society's point of view because it restricts uses of the new product and/or process (relative to the competitive provision of the innovation). Earlier economic analysis focused extensively on this issue and questioned the economic desirability of the patent system (Machlup and Penrose 1950). But the profit opportunity created by the monopolistic control of the innovation can be a powerful *ex ante* incentive, presumably enough to motivate R&D investments that would otherwise be neglected.

Patents Can Help the Dissemination of Knowledge

An additional benefit of patents is related to the disclosure requirement. In most countries, patents are disclosed 18 months after the filing date or earlier. The importance of this feature is predicated on the fact that, absent patents, inventors can rely on trade secrets to protect their discoveries (Friedman, Landes, and Posner 1991). By providing an incentive for disclosure, patents are held to contribute to a desirable dissemination of scientific and technical information, allowing other inventors to avoid duplicating existing discoveries and making it easier to develop further innovations that build on the known state of the arts (possibly by "inventing around" a patent as well). It is useful to note that the disclosure argument offers an economic role for patents, even for inventions that have already taken place, and as such it is quite distinct from the incentive role of patents due to the increased appropriability of R&D output.

Patents Can Avoid Wasteful Innovation Efforts

As the arguments on the disclosure property of patents suggest, an important and beneficial effect of new knowledge is that it makes possible further innovations and discoveries. Discoveries from basic research are often of this sort, at times opening up entire new fields of research. It can be argued that patenting of such seminal inventions can have useful social payoffs. This rationale is articulated in the so-called prospect theory of patents, originated by Kitch (1977). It relies on the notion that broad, early property rights on key inventions allow an orderly pursuit of follow-up innovations and reduce wasteful innovation races. An analogy can be made to the practice of granting mineral claims on land where no discovery has yet been made, to avoid a wasteful mining of the prospect. Whereas patents in such cases can clearly have positive efficiency effects, it is also easy to see that broad, early patents can adversely affect further research, especially when the original discovery has applicability in many uses. If the original inventor does not have a comparative research advantage or interest in pursuing some research directions, and licensing of the patented innovation to third parties is problematic (perhaps because of excessive transaction costs), patenting can have adverse effects on the flow of further innovations. We will return to related issues later when discussing the appropriate breadth of patents.

Patents Can Help Technology Transfer and Commercialization

Whereas the difficulty of licensing may reduce the desirability of exceedingly broad property rights, patents can actually play a critical role in licensing and, more generally, in the dissemination of new knowledge. A key insight here is due to Arrow (1962), who stressed the information nature of new knowledge and noted a peculiar property in the determination of demand for information. Specifically, to put a value on information a would-be purchaser needs to have the information, but at that point, if the seller does not have property rights on the information, the would-be buyer has no more incentive to pay for it. Patents, therefore, can play a crucial role in reducing transaction costs of licensing innovation and, more generally, in technology transfer.

A related but distinct role for patents as instruments of technology transfer has been articulated to rationalize the Patent and Trademark Laws Amendments of 1980, commonly known as the Bayh-Dole Act. The main elements of this reform were to allow universities and other entities to patent, retain title to, and commercialize federally funded inventions; and, to allow federal agencies to grant exclusive licensing for their inventions. Based on the view that the main role of patents is to provide incentives for innovation that would not occur otherwise, it would be difficult to make an economic case for public institutions patenting discoveries that already have been publicly funded and accomplished. Likewise, the role of patents in transferring information would be irrelevant in this case, because public research institutions have little use for trade secrets, and because it is difficult to improve on the dissemination of information achieved by simply publishing a discovery. But the presumption here is that many discoveries produced by publicly funded R&D, and in the public domain, may not be used in technological developments because, without an exclusive license backed up by patent rights, firms would not be interested in expensive development work required to transform an invention into a new product. Little evidence exists to corroborate this belief for the case of university patenting (Mowery et al. 2001). But more generally, when it is difficult to assert property rights on development work, a patent on underlying innovations obviously may aid development and commercialization of new technologies.

Length and Scope of Patent Protection

As illustrated in the foregoing, the patent system entails a fundamental trade-off—inventors are given a monopoly position (which entails inefficiencies) in order to provide them with incentives to innovate (which carries economic benefits).[3] In other words, the dynamic efficiency of encouraging invention leads to static inefficiency created by a monopolistic situation. Given that, a natural question to ask is, what is the optimal degree of patent protection? The degree of market power provided by a patent essentially depends on three elements: the length (duration) of the patent, the breadth of the patent, and the height of the patent. We analyze these features of patent protection for the case of a single innovation first, and consider the issues related to cumulative innovations later on.

Duration of the Monopoly Power

The length of the patent determines the duration of monopoly power that a patentee can expect and thus affects the static inefficiency as well as the incentive role of patents. The benefit to the innovator increases with the length of the patent. But from society's point of view, too short a patent may dissuade research, whereas a patent that is too long may give excessive rent to the owner and may block further improvement.

Consider first a given innovation, such as the new product case illustrated in Figure 1, where Π denotes the per-period profit accruing to the patentee. Here, the optimal duration of the patent should be determined so as to recoup exactly the R&D expenditure necessary to bring about the innovation. Then, for a discount rate r, and given a one-time research cost F of bringing about the innovation, the optimal patent length T^* would solve $\int_0^{T^*} \Pi e^{-rt} dt = F$. A more realistic model would allow the size of the innovation to vary with the investment in R&D, which in turn depends on the length of the patent. For simplicity, think of a process innovation that leads to a reduction in the unit cost of producing a good. The optimal patent length must now balance the social gain from the innovation due to a larger reduction in cost and the social loss associated with a longer delay in the exploitation of the innovation by rival firms. This point was originally made by Nordhaus (1969), who found that an optimal patent must be of finite duration but strictly positive, as shown later by Scherer (1972).

In these earlier models, it was assumed that R&D investment always leads to an innovation the size of which depends on the investment undertaken and where an R&D project necessarily succeeds. This is a restrictive assumption that does not take into account competition at the R&D level. The introduction of such competition can, under certain circumstances, increase the optimal length. Indeed, the expected payoff of each competitor is reduced; therefore, for a given number of competitors, an increase in the patent duration may be the only way to reestablish the R&D incentive (Kamien and Schwartz 1974). But in other circumstances, competition can reduce the required patent length, as is the case in DeBrock's (1985) model where it is assumed that competition takes place at the research level and not at the development level where patent holders can develop their innovation without threat of competition. Here, R&D investment at the development stage is the same as without competition, but competition at the research level leads to useless duplication of costs for firms that cannot get a patent, and that in turn reduces the social surplus.

The industrial context may play an important role in the determination of the optimal patent length. For instance, in markets where demand is very elastic (i.e., quantity demanded is very responsive to a small change in price), monopolistic pricing of the innovation can lead to large *ex post* welfare losses (i.e., a large L in Figure 1). Hence, in such markets a shorter patent length would be warranted. The optimal patent length is also likely to be innovation-specific, and thus the one-size-fits-all patent length (20 years from filing) is sub-optimal. Roughly speaking, a uniform patent length provides too much protection for "easy" innovations (those that would have been pursued even with a shorter patent period), thereby creating unnecessary efficiency losses, whereas it provides too little protection for "difficult" innovations, such that some research projects that are socially desirable do not get undertaken.[4]

The Scope of a Patent

Whereas the length of the patent protection characterizes the duration of the monopoly power, the scope of a patent bears on the intensity of the induced monopoly power (Merges and Nelson 1990). The breadth of a patent defines the range of products that are encompassed by the claims of the patent and therefore protects the patent holder against potential imitators. In general, the less specific the claims of the patent are, the broader the patent. The height of a patent, on the other hand, confers protection against improvements or applications that are easy or trivial. The value of a patent to a firm depends on how effective its protection is in these two dimensions (breadth and height), in addition to being related monotonically to the patent length.

Unlike its maximum length, which is fixed by law, patent breadth is, to a certain extent, endogenous. Patent breadth depends on the claims put forth by the patentee and also is a feature over which the patent office has some discretion (at the examination stage).[5] Clearly, a reduction in the breadth of patents would induce more competition (e.g., imitation), which benefits consumers. But too narrow a patent reduces the incentive to innovate. What is the optimal breadth for patents? Economic analysis suggests two kinds of results in this setting. Narrow and long patents can be found to be optimal because broad patents are costly for society in that they give excessive monopoly power to the patent holder (Gilbert and Shapiro 1990). Central to this conclusion is that the flow of payoff from holding a patent has a negative impact on the social surplus. Thus, a minimum level of flow of payoff (breadth), with duration adjusted accordingly, would be socially optimal. In Klemperer's (1990) more general model, both narrow and long, or broad and short, patents can be optimal, depending on the structure of demand. Broad and short patents can also be optimal when they discourage imitation and thus enhance the incentive to innovate (Gallini 1992). Specifically, imitation is discouraged when it is too costly (broad patent) and when imitators do not have enough time to enter the market (short patent).

Denicolò (1996) analyzes these alternative conditions in a single model in which the problem is to minimize a ratio of social loss to the incentive to innovate. The breadth influences both the elements of the ratio, and, depending on the assumptions about the nature of competition in these markets, all the above results can be found. Thus, in different industries, different optimal patent policy could be enforced. Patents in biotechnology should be different from patents in e-commerce, for example. Note that in these models, patent length and the scope of protection are substitutes for the purpose of providing a given level of protection to the innovator. This is not necessarily the case when we consider cumulative innovations, as we will see below.

Another dimension of the patent right is the protection it gives to the innovator against improvements that are too close to the patented innovation. This feature is sometimes referred to as the novelty requirement, also called “height” (van Dijk 1992) or “leading breadth” (O’Donoghue, Scotchmer, and Thisse 1998). Competition from follow-up innovations are clearly affected by the height of the patent, and too high a patent gives excessive monopoly power to the patent holder. As for the interaction between height and duration, La Manna (1992) shows that a patent of infinite duration and finite height can be optimal.

In conclusion, patent scope (breadth and height) can substitute for patent length in providing *ex post* returns from (and therefore *ex ante* incentive for) an innovation. But one should not conclude that broader or higher patents necessarily induce more research. The problem is that, *ex ante*, a potential inventor needs to consider the possibility of discovering something that may infringe on existing patents. The broader or higher the patents, the greater the risk of such a possible unwanted infringement, and this possibility can discourage R&D investments. As discussed in the foregoing, specific conclusions about optimal patents are sensitive to the way one chooses to model the innovation process. Also, the degree of substitutability between length, breadth, and height of a patent depends on the nature of the innovation. Currently, there seems to be widespread concern about exceedingly broad patents being granted, particularly in recent innovation fields such as biotechnology and so-called business methods (Merges 1999). This is particularly the case for cumulative innovations, when an innovation is useful mostly as a research tool in other R&D activities.

Cumulative Innovations

The importance of patent height is most apparent for the case of sequential innovations. Such innovations are particularly relevant when the innovation is a research tool, used mainly in the R&D process for further innovations. Consider first the case of a basic innovation followed by its improvement, which eventually can be accomplished by the same innovator. What is the impact of the scope of the protection in the diffusion of innovations over time? Scotchmer and Green (1990) focus on the profit incentive for R&D when the second innovation is more profitable than the first one (from social and private viewpoints) and when there exists an information externality among innovators (patent disclosure confers a positive externality on the innovator's competitors). The policy question here is to determine how stringent the novelty requirement should be in order to protect the profit of innovators while encouraging disclosure of innovation. It is presumed that the reasons for granting a patent are to create an incentive to do research and to accelerate aggregate innovation through disclosure of innovations. The first reason calls for a strong novelty requirement (any small improvement should be found to infringe upon the patented innovation and thus protect fully the first innovator), whereas the second reason calls for a weak novelty requirement.

Because the information externality that the innovator may want to avoid is due to the disclosure requirement of patents, the innovator may prefer to keep small innovations secret. The strategic suppression of small improvements can avoid the *ex ante* erosion of profit due to a weak novelty requirement (Scotchmer and Green 1990). The weak novelty requirement seems to be socially preferable because an early disclosure accelerates the introduction of the improvement. Nevertheless, even in this system, the first innovation is not always patented. When the *ex ante* profit is larger with the strong novelty requirement, this last system is socially preferable. In any case, the first innovator can choose not to patent the innovation. One way to force firms to disclose their small innovations without undermining their profit would be to force them to cooperate. This could take the form of licensing (Green and Scotchmer 1995; Scotchmer 1991). In the same vein, Matutes, Regibeau, and Rockett (1996) show that a system based on an optimal height rather than an optimal length is socially preferable because it induces an earlier introduction of the innovation.

However, it is not always the case that the first innovator is able to improve on her innovation or even to develop applications. For instance, small laboratories that do not have access to sufficient capital may be unable to perform further development. In this setting, what is the appropriate patent scope that gives enough incentive to the first and second innovators to undertake R&D investment? Let us first assume that only one firm is capable of being the second innovator. In this setting, each innovator should receive an incentive at the level of the social benefit she creates. For the second innovator, it should be the incremental benefit created by her improvement, whereas for the first innovator, it should be the sum of the first innovation benefit and the incremental benefit. Indeed, the feasibility of the second innovation is due to the very existence of the first one. For this reason, the actual profit-sharing arrangement between innovators is important. If profit sharing is in favor of the first innovator it will promote radical innovations, whereas if it is in favor of the second innovator it will promote improvements.

It is nevertheless very difficult to favor improvement if the first innovation has no value by itself (but is fundamental to the discovery of the second innovation). On the other hand, a policy in favor of the first innovator will reduce follow-up improvements and will thus affect the profit of the first innovator. Public authorities have two ways to intervene in the profit-sharing process: through the scope of the patent (height, length, or patentability of the second generation of innovations), and through the competitive policy in regard to licenses. Green and Scotchmer (1995) present a situation in which it is profitable for innovators to sign *ex ante* license agreements. They show that an infinite height (i.e., any second-generation innovators need to get a license from the original innovator) is optimal when there is certainty about the return of the second generation of innovations. In the case of uncertainty, that is no longer the case.

Scotchmer (1991) explores the role of the patentability of the second generation of innovations when the improvement of the initial innovation can be done by more than one firm. The very existence of competition to obtain the second patent reduces the expected profit of the first innovator and thus reduces the attractiveness of the *ex ante* license. An *ex ante* license agreement can be impossible to implement in situations where the second generation of innovators are reluctant to disclose their ideas for improvement to the first innovator. Chang (1995) considers this situation and shows that, no matter what the value of the first innovation, there exists a threshold value for the second innovation below which a license is required and above which it is not required. The smaller the benefit of the first innovation, the higher must be the patent in order to provide enough incentive to the first innovator. But optimal patent height is not a monotonic function of the social value of the first innovation, and when this value is large, the patent also must be high.

To fully capture the impact of the cumulative nature of the innovations on the incentive to innovate, it is useful to consider a dynamic model where improvements of the innovation arise randomly (O'Donoghue, Scotchmer, and Thisse 1998). In this setting, the innovators must be protected against improvements (leading breadth or height) and against imitations (lagging breadth or breadth). Here it is assumed that the probability that the same innovator makes two successive improvements is almost nil. Depending on the rate at which improvement ideas arise, under a perfect lagging breadth, innovators may overinvest or underinvest. They will tend to underinvest if ideas are too frequent because they fear that another improvement will be introduced too early. But if ideas are not that frequent, firms will overinvest because each innovator expects to increase her payoff once she becomes a follower after having been a leader. With a maximal lagging breadth, and if leading breadth is finite but duration is infinite, firms underinvest in R&D. An infinite duration for a low patent reduces the cost associated with the delay in diffusion (O'Donoghue, Scotchmer and Thisse 1998).

Thus, when innovation is cumulative, the acceleration of the disclosure of the first innovation is crucial to allow further improvements. However, this must be done without altering the incentives of the first innovator. Furthermore, when innovations are cumulative, their complementarity gives rise to additional issues.

Complementary Innovations

To manufacture a complex new product (e.g., a computer chip or a transgenic crop), it is common for firms to need a number of intermediate inputs that are patented, with such patent rights likely held by different firms. Insofar as the patented inputs are highly complementary, or even essential, in the manufacture of the new product, the suppliers of these inputs essentially hold blocking patents. The danger then is that the manufacturer becomes susceptible to *hold-up* by the patentees (Shapiro 2001). In the context of biotechnology innovations, this problem has been characterized as the "tragedy of the anticommons" (Heller and Eisenberg 1998). Recall that the standard notion of tragedy of the commons refers to the overuse of a common property resource, and it arises because not enough property rights are allocated. The tragedy of the anticommons situation arises because, somewhat paradoxically, there is an excess allocation of property rights (i.e., too many gatekeepers with the right to levy a tax), and this situation can lead to an underuse of the resource (in this case, the pool of knowledge).

The hold-up problem when patented innovations are complements can be quite damaging to the functioning of the patent system. The issue here is mostly one of coordination among patent holders. Mechanisms that can address the hold-up problem in this setting include cross-licensing and patent pools. With cross-licensing, which can be implemented with or without license fees, two (or more) firms agree to share a subset of each other's patented technologies. This kind of arrangement is particularly attractive when the relevant patents are held by the manufacturers of the new product that needs the complementary patented inputs. Cross-licensing is apparently quite common between firms engaged in the design and manufacture of microprocessors, for example. A patent pool is an arrangement whereby a set of patents is licensed in a single package, for a posted fee, either by a single owner or by an entity especially set up to handle this arrangement. The obvious danger here is that, whereas the inclusion of complementary patents in the pool is justifiable (as detailed in the foregoing discussion) and leads to pro-competitive behavior, the inclusion of substitute (rival) patents in the pool could be a means to implement collusion and could exacerbate noncompetitive behavior. Patent pools, therefore, can pose antitrust issues. Another way to reduce the fragmentation of patent ownership is through mergers and acquisitions, and that, indeed, appears to be one of the driving forces behind the recent industry consolidation in the life sciences sector.

Patents, Market Structure, and the Incentive for R&D

The research activity of innovators undoubtedly is influenced by the patent system, as well as by the industry structure. To gain some insights into the interaction of patents and industry structure, we explore how innovators invest in R&D. Note first that the discovery of an innovation affects the structure of the market. This is easy to understand if we consider a process innovation that allows reducing production costs: firms can compete in the pre-innovation and, if a patent is granted, the holder will be a monopolist in the post-innovation market. Competition can also take place at the R&D level. The notion that firms may be competing at the innovation stage has been a recurring one in this paper. This competition is very much affected by the patent system, which has the feature of a winner-takes-all contest. To analyze the effects that this may have on R&D, recall that the presumption is that, absent patents, economic agents do not have sufficient incentives to invest in R&D activities and, consequently, not enough innovation activity would take place. The term of comparison here is usually the "socially optimal" level of R&D activities, which one can define (at least conceptually) taking into account the overall expected benefits brought about by an innovation, as well as the costs of R&D projects. This first-best solution typically cannot be achieved by a decentralized market system, for lack of incentives and coordination mechanisms (the "market failure" predicament), and it cannot be achieved by a centralized government either, for lack of information and implementation mechanisms. Given these benchmarks, the patent system is likely to increase private investments towards the first-best (socially optimal) level of R&D.

To isolate the pure incentive to innovate (that is, without any competition at the innovation stage), it is necessary to note that the value of an innovation to an inventor depends on the nature of the innovation but also on the market structure that applies at the production stage (Arrow 1962). Monopoly and competition provide useful polar cases. It turns out that, *ceteris paribus*, the incentive to innovate is lower in a monopolistic market than in a competitive market. This is due to the replacement effect: a monopolist has less to gain from a given innovation (say, a cost-reducing innovation) because one monopolistic situation replaces another monopolistic situation. But even under competitive production conditions the incentive to innovate is less than what is socially optimal. This latter case is actually depicted in Figure 1, which can be interpreted as representing a market where competitive producers with unit cost c must pay a royalty to the inventor. In this situation it is optimal for the inventor to set the royalty rate at $(p^M - c)$, and thus the incentive to innovate is the profit Π. However, from society's perspective, the incentive to innovate ought to be the entire (potential) consumer surplus $(S + \Pi + L)$.

What is missing from the foregoing analysis is the possibility of entry in the innovation industry; typically it is possible (and likely) for more than one firm to pursue the same research result. The resulting "patent race" tends to dissipate the expected rent from patenting and to increase aggregate R&D investment (in a fashion akin to the "common pool" problem of competitive fishing in an open access fishery). To see what some of the implications might be, following Wright (1983) and Tandon (1983), suppose that n firms pursue the same invention. Each firm incurs a one-time research cost F. Let $P(n)$ represent the probability that the invention takes place (at least one firm is successful). In general, more firms increase the aggregate success probability but at a decreasing rate (i.e., $P'(n) > 0$ and $P''(n) < 0$). For simplicity, assume that the social benefits from the discovery are equal to the private benefits that accrue to the inventor, say B. From a social point of view, the optimal number of firms n^* that pursue the invention should maximize the expected net social benefit and thus should solve $P'(n^*)B = F$ (i.e., expected marginal social benefit = marginal social cost). But because each individual firm has an incentive to join the patent race as long as the expected profit is greater than the cost F (under risk neutrality), the equilibrium number of firms n^C in competition solves $P(n^C)B/n^C = F$. Given the properties of the aggregate probability function $P(n)$, it then must be that $n^C > n^*$, and therefore excessive resources are allocated to R&D by this patent race.

Consider now the dimension of time and suppose that the date of discovery of the invention is a decreasing function with the investment in research (Dasgupta and Stiglitz 1980), and assume that the investment is completely paid at date zero. The pre-innovation situation is competitive, and only the firm that first makes the innovation will be granted a patent (monopolistic post-innovation situation). This situation is similar to an auction in which at date zero the patent will be granted to the firm that proposes to innovate first. In this setting, only one firm will invest exactly the amount that it will gain from innovating. The deterministic nature of the relationship between the date of discovery and the investment makes it easy to find that the investment undertaken by a monopolist is smaller than the investment undertaken by competitive firms. Furthermore, the monopolist invests less than that which is socially optimal. This is due to the negative externality of the investment of an innovator on her rivals, which none of the firms internalize.

In this setting, what happens to the incentives to innovate if a potential entrant can introduce the innovation instead of the incumbent? In other words, do technological innovations come from a monopolist in the pre-innovation market or from new innovators? The pure incentive to invest is higher for potential entrants, due to the replacement effect. But there is another effect that has to be considered: the efficiency effect that arises because the profit of a monopoly is bigger than the sum of the duopoly profits. Thus, a monopolist has more to lose when a competitor enters the post-innovation market compared to a potential entrant. This leads to preemption (Gilbert and Newbery 1982) and to the persistence of monopoly.

Now we add one more degree of realism and consider not only that discovery depends on the R&D investment but also that it is random, as in the case of "patent race" models with stochastic innovation (Loury, 1979; Lee and Wilde 1980; Reinganum 1989). The date of discovery is no longer deterministic but random according to a Poisson process. The investment can be fixed and incurred at the beginning (Loury 1979), or it can be incurred as a constant flow over time (Lee and Wilde 1980). The advantage of the second model is that firms can stop to invest as soon as a rival has made a discovery. These models have no memory: the probability of success does not vary as time goes by or as investment is accumulated. Each firm considers only its own marginal benefit and does not take into account the reduction that its own investment will impose on its rival's payoff. It is not known when a discovery will be made (technological uncertainty), nor who will make it (market uncertainly). Loury (1979) shows that as the number of firms that compete increases, the average optimal investment decreases. On the contrary, in the model of Lee and Wilde (1980), it increases. The difference in results is due to the different structure of investment. Nevertheless, for a given number of firms engaged in the race, firms tend to overinvest in R&D compared to the socially optimal level in both models. Thus, there is a tendency for excessive research.

In this setting, do monopolists have a stronger incentive to invest than do potential entrants? Which effect (replacement or efficiency) dominates the other depends on the nature of the innovation. For an innovation that is almost drastic, a monopolist invests less than a potential entrant and thus preemption disappears (Reinganum 1983).[6] There is no efficiency effect anymore; only the replacement effect plays a role. When the firm behind in the race can leapfrog, or when the process is without memory (i.e., at each point in time the probabilities of winning are the same for each firm), there is no preemption (Fudenberg et al. 1983). On the other hand, when it is not possible to leapfrog, as is the case in Gilbert and Newbery (1982), then there is always persistence of monopoly.

Thus, firms engaged in a patent race may collectively invest more than that which is socially desirable. Furthermore, firms engaged in a patent race are likely to choose research strategies that are too risky. Because the race is essentially a winner-takes-all contest, there is a clear payoff to being first, but the penalty for losing does not depend on how far behind a firm ends up. And firms will tend to choose research projects that, from a social perspective, display too much correlation (i.e., there is too much duplication of research efforts) (Dasgupta and Maskin 1987). To illustrate this last point, suppose that it is socially desirable for a particular firm to choose a research direction very different from that of its competitors (because that would increase the probability of someone being successful). The reason an individual firm may not want to follow such a (socially optimal) strategy is because by doing so it would increase the probability that a competitor is successful in instances when the firm is not.

Beyond Patents

The main justification for patents is to foster innovation in a market economy, but the patent system is not the only method for encouraging innovation. Copyrights and trademarks, of course, are additional instruments for intellectual property right protection that typically apply in contexts where patents do not. Trade secrets, on the other hand, can apply to patentable innovations and can provide effective protection against another party's discovery by inappropriate means (although a trade secret offers no protection against independent discovery or reverse engineering). Certain biological innovations are also afforded *sui generis* protection by such means as plant patents and plant variety protection certificates (Barton 1998).

More generally, alternatives to patents include rewards or prizes, procurement contracts, and public production of new knowledge. With the reward system, the government specifies a fixed sum of money for a well-defined research goal and then awards this "prize" to the first firm to achieve the desired result. Asymmetric information between researchers and the government can make it difficult to implement the reward mechanism (Wright 1983). Specifically, to be effective, the government must know about the feasibility of various research projects as well as be able to assess the demand for various potential innovations. But firms are likely to be better informed than the government on such matters, and a decentralized solution such as the patent system may be superior.

With the procurement system, the government picks the firms that will be involved in the research project and specifies the terms of the project (such as expected research output and compensation terms) in a binding contract (Laffont and Tirole 1993). Unlike the prize system, this method can eliminate unwanted duplication of research efforts. But again, for this system to be efficient, the government must be quite knowledgeable about the costs and benefits of research ventures.[7] More generally, innovations can be stimulated by the government's direct involvement in the production of new knowledge. Much of the research carried out at public institutions, and sponsored by public funds, is an example of such knowledge production. This structure, together with the complex social milieu characterizing academic institutions, has made possible the "open science" environment that can take substantial credit for many scientific and technological breakthroughs (Stephan 1996; David 1998). The tension between the behavioral standards of open science and the privatization of new knowledge made possible by patents is readily apparent. The concern is that the increased reliance on intellectual property right mechanisms may be eroding the domain of public information and access to research tools, which could have serious long-term consequences for the vitality of the community of science (David 2000).

Conclusion

The patent system has emerged as the central institution for asserting intellectual property rights in many crucial fields of science and technology. From an economic point of view, patents offer a second-best solution to the market failure arising from the public-good nature of knowledge. As such, the patent system contributes to solving a problem but comes with shortcomings of its own, mostly because it creates market power positions that can adversely affect the economic performance of the system. In fact, for most of the nineteenth century, the patent system was under considerable criticism by the same economics profession that now provides the most valuable insights for its defense. This change is due to the increased appreciation for the critical role that innovations play in stimulating economic growth. The possibility of protecting discoveries through patents, for example, is credited for bringing about crucial technical improvements in the industrial revolution (Dutton 1984).

As noted, the *ex post* inefficiency of the patent system is viewed as the necessary downside in providing enough inducement to undertake desirable R&D projects. The size of the inducement depends on the length and scope of the patent right. Ideally, such an inducement should be proportional to the cost of the R&D project, which means that the length, breadth, and height of a patent should be tailored to each particular innovation. In addition to the cost of the project, such a tailored patent should also reflect the particular market conditions of the new product and/or process. Clearly, the patent system does not do that, and arguably it cannot do that. These limitations suggest that continued efforts are required to improve the workings of the patent system. A solid understanding of its complex (and sometime subtle) economic implications, which we have tried to review here, should prove useful in this endeavor.

Endnotes

1. Although this body of work reflects accepted stylized facts and empirical findings, as well as the evolving legal and institutional features of the patent system, there is also a distinct and large set of empirical studies on the economics of patents that we will not review here.

2. Here, S measures the consumer surplus while the patent is active, and $\left(S+\Pi+L\right)$ is the consumer surplus after the patent has expired. Thus, L is the "loss" to society from underutilizing the innovation during the life of the patent.

3. Here and in what follows, we do not distinguish between the incentive to create and the incentive to commercialize an innovation.

4. These considerations are somewhat tempered by the fact that effectively different patent lengths can be implemented through the imposition of different renewal fees (Cornelli and Schankerman 1999). In most European countries, patent holders must pay an annual renewal fee in order to keep their patents in force. In the United States, only three renewal fees have to be paid: at patent ages 3, 7, and 11.

5. In U.S. case law, breadth and height of a patent also are upheld by the so-called doctrine of equivalents, which asserts that a new product that is outside the stated claims of an existing patent could still infringe on this patent if it is essentially an equivalent product.

6. As in Arrow (1962), an innovation is "drastic" if its pricing is not affected by the threat of competition (the innovator can behave as an unrestricted monopolist).

7. This issue may be less problematic when the government is the only intended customer of the innovation, as happens, for example, with research related to national defense.

References

Arrow, K.J. 1962. "Economic Welfare and the Allocation of Resources for Inventions." In *The Rate and Direction of Inventive Activity: Economic and Social Factors.* Edited by R.R. Nelson. Princeton, NJ: Princeton University Press.

Barton, J.H. 1998. "Acquiring Protection for Improved Germplasm and Inbred Lines." In *Intellectual Property Rights in Agricultural Biotechnology*. Edited by F.H. Erbisch and K.M. Maredia. Oxon, UK: CAB International.

Barzel, Y. 1989. *Economic Analysis of Property Rights*, 2nd edition. Cambridge, UK: Cambridge University Press.

Chang, H. 1995. "Patent Scope, Antitrust Policy and Cumulative Innovation." *RAND Journal of Economics* 26: 34-57.

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

David, P. 1998. "Common Agency and the Emergence of 'Open Science' Institutions." *American Economic Review—Papers and Proceedings* 88: 15-21.

———. 2000. "A Tragedy of the Public Knowledge 'Commons'? Global Science, Intellectual Property and the Digital Technology Boomerang?" Stanford Institute for Economic Policy Research, Discussion Paper No. 00-02, Stanford University, October.

Dasgupta, P., and E. Maskin. 1987. "The Simple Economics of Research Portfolios." *Economic Journal* 97: 581-95.

Dasgupta, P., and J. Stiglitz. 1980. "Uncertainty, Industrial Structure and the Speed of R&D." *Bell Journal of Economics* 1: 1-28.

DeBrock, L.M. 1985. "Market Structure, Innovation and Optimal Patent Life." *Journal of Law and Economics* 28: 223-44.

Denicolò, V. 1996. "Patent Race and the Optimal Patent Breadth and Length." *Journal of Industrial Economics* 44: 249-65.

Dutton, H. 1984. *The Patent System and Inventive Activities during the Industrial Revolution 1750-1852*. Manchester, UK: Manchester University Press.

Friedman, D.D., W.M. Landes, and R.A. Posner. 1991. "Some Economics of Trade Secret Law." *Journal of Economic Perspectives* 5: 61-72.

Fudenberg, D., R. Gilbert, J. Stiglitz, and J. Tirole. 1983. "Preemptive, Leapfrogging and Competition in Patent Race." *European Economic Review* 22: 3-31.

Gallini, N. 1992. "Patent Policy and Costly Imitation." *RAND Journal of Economics* 23: 52-63.

Gilbert, R., and D. Newbery. 1982. "Preemptive Patenting and the Persistence of Monopoly." *American Economic Review* 72: 514-26.

Gilbert, R., and C. Shapiro. 1990. "Optimal Patent Length and Breadth." *RAND Journal of Economics* 21: 106-12.

Green, J.R., and S. Scotchmer. 1995. "On the Division of Profit in Sequential Innovation." *The RAND Journal of Economics* 26(1): 20-33.

Heller, M.A., and R.S. Eisenberg. 1998. "Can Patents Deter Innovations? The Anticommons in Biomedical Research." *Science* 280: 698-701.

Kamien, M.I., and N.L. Schwartz. 1974. "Patent Life and R&D Rivalry." *American Economic Review* 64: 183-87.

Kitch, E.W. 1977. "The Nature and Function of the Patent System." *Journal of Law and Economics* 20: 265-90.

Klemperer, P. 1990. "How Broad Should the Scope of Patent Protection Be?" *RAND Journal of Economics* 21: 113-30.

Laffont, J-J., and J. Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

La Manna, M. 1992. "Optimal Patent Life vs. Optimal Patentability Standard." *International Journal of Industrial Organization* 10: 81-90.

Lee, T., and L. Wilde. 1980. "Market Structure and Innovation: A Reformulation." *Quarterly Journal of Economics* 94: 429-36.

Loury, G. 1979. "Market Structure and Innovation." *Quarterly Journal of Economics* 93: 395-405.

Machlup, F., and E. Penrose. 1950. "The Patent Controversy in the Nineteenth Century." *Journal of Economic History* 10(May): 1-29.

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

Mazzoleni, R., and R.R. Nelson. 1998. "The Benefits and Costs of Strong Patent Protection: A Contribution to the Current Debate." *Research Policy* 27: 273-84.

Merges, R.P. 1997. *Patent Law and Policy—Cases and Materials*, 2nd edition. Charlottesville, VA: Michie Law Publishers.

———. 1999. "As Many as Six Impossible Patents before Breakfast: Property Rights for Business Concepts and Patent System Reform." *Berkeley Technology Law Journal* 14: 577-615.

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

Mowery, D.C., R.R. Nelson, B.N. Sampat, and A.A. Ziedonis. 2001. "The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980." *Research Policy* 30: 99-119.

Nordhaus, W.D. 1969. *Inventions, Growth and Welfare: A Theoretical Treatment of Technological Change.* Cambridge, MA: MIT Press.

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

Reinganum, N. 1983. "Uncertain Innovation and the Persistence of Monopoly." *American Economic Review* 73: 741-48.

———. 1989. "The Timing of Innovation." In *Handbook of Industrial Organization*, vol. 1. Edited by R. Schmalensee and R.D. Willig. Amsterdam: North-Holland, 849-908.

Scherer, F.M. 1972. "Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation." *American Economic Review* 62: 422-27.

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

Scotchmer, S., and J. Green. 1990. "Novelty and Disclosure in Patent Law." *RAND Journal of Economics* 21: 131-46.

Shapiro, C. 2001. "Navigating the Patent Thicket: Cross-licenses, Patent Pools, and Standard-Setting." In *Innovation Policy and the Economy*, vol. 1. Edited by A. Jaffe, J. Lerner, and S. Stern. Cambridge, MA: MIT Press.

Stephan, P.E. 1996. "The Economics of Science." *Journal of Economic Literature* 34: 1199-1235.

Tandon, P. 1983. "Rivalry and the Excessive Allocation of Resources to Research." *Bell Journal of Economics* 14: 152-65.

van Dijk, T. 1996. "Patent Height and Competition in Product Improvements." *Journal of Industrial Economics* 44: 151-67.

Wright, B. 1983. "The Economics of Invention Incentives: Patents, Prizes, and Incentive Contracts." *American Economic Review* 73: 691-707.