REFGOV

Reflexive Governance in the Public Interest

Institutional Frames for Markets

The Properties of Intellectual Property

By Antonio Nicita and Maria Alessandra Rossi

Working paper series: REFGOV-IFM-92
The Properties of Intellectual Property

Antonio Nicita
Maria Alessandra Rossi

University of Siena

paper prepared for the VIth EU Framework Project
"Reflexive Governance in the Public Interest"

Project no CIT3-CT-2005-513420

REFGOV
Reflexive Governance in the Public Interest

WP15 - Intellectual Property Rights, Incentives to Invent, to Accumulate Knowledge and to Circulate Intangibles (WP15)
Table of Contents

1. Introduction ..........................................................................................................................5
2. The rationale for providing incentives to innovative activity ...........................7
3. IP put in context: incentive mechanisms alternative to IP .................................9
4. Intellectual property rights: the legal perspective .....................................................15
   4.1. Utilitarianism ..............................................................................................................15
   4.2. Labor Theory .............................................................................................................18
   4.3. Personality Theory .....................................................................................................22
   4.4. Social Planning Theory .............................................................................................23
5. Intellectual property rights: the economic perspective ...........................................24
   5.1. The economic functions of the patent system .......................................................24
       5.1.1. IP incentive function ......................................................................................24
       5.1.2. IP transactional function ...............................................................................26
       5.1.3. IP disclosure function ...................................................................................28
       5.1.4. IP signalling function ....................................................................................31
   5.2. The optimal design of the patent system .................................................................32
   5.3. The role of IP enforcement and litigation ...............................................................38
       5.3.1. Empirical evidence on patent litigation .........................................................40
       5.3.2. Theoretical analyses of patent litigation .........................................................42
6. Concluding remarks ..........................................................................................................45
1. Introduction

The need for publicly provided incentives to innovation appears to be widely acknowledged in present-day economies. The production of innovative knowledge is not only understood as a crucial aspect of competition among private firms, but is increasingly recognized as the primary source of countries’ comparative advantage. In the past, some critical voices have stood against any form of government action in support of innovation, on the grounds that it is superfluous (Plant, 1934; Breyer, 1970). By contrast, current scholarship and policymaking is more concerned with the question of what form public incentives to innovate should take rather than whether they should be provided at all.

Perhaps the most controversial among the forms of publicly provided incentives to innovate is intellectual property (in what follows, also IP). Throughout the centuries intellectual property has been the subject of heated discussion. Many writers have contributed to the shaping of this highly complex area, which involves elements of economics, law, international relations, politics and ethics.

In the 19th century intellectual property rights, or more precisely patents, started gripping public imagination at least as much as they ignited theoretical speculation, as public attention was cast on folk heroes such as Thomas Edison and crowds gathered to see his inventions. At present, even though public opinion is once again to some extent mobilized over intellectual property issues (the recent European debate over software patentability is just the most prominent example but by no means the only one), scholarly disputes tend to disregard the ethical concerns that were the subject of Thomas Jefferson’s or Karl Marx’s discussions of IP, in favour of a more pragmatic approach. Nowadays, the debate over intellectual property is mainly one about the choice of the set of public policy instruments most suitable to the enhancement of technological progress.

The reason why intellectual property has been trusted back into the spotlight
relates to the emergence of a number of concerns over the restrictions to access to intangible resources and the restraint of development potentially associated to it. In particular, the following developments are particularly salient: (a) the tendency of patents to protect knowledge resources ever closer to the realm of ideas and pure knowledge; (b) the ever less stringent application of patentability criteria (Barton, 2000; Lunney, 2001); (c) the tendency to adopt restrictive contractual practices in the management of intellectual resources (refusals to license, exclusive licensing, “reach-through claims” e grant-back clauses; OECD, 2002) and to make recourse to the strategic use of proprietary intellectual assets also through the use of so-called patent trolls (Cohen et al., 2000; Rivette and Kline, 2000); (d) the increasing recourse to patenting for the results of public research, promoted by the Bayh-Dole Act in the US in 1980 and subsequently by similar legislative interventions in other countries (Rai and Eisenberg, 2003); and (e) the ever more subtle distinction between basic research and applied research, that undermines the willingness of private actors to grant privileged access to proprietary intellectual resources on the basis of a sort of division of labour between the public and private sphere that parallels the distinction between basic and applied research (Cooper Dreyfuss, 2004).

The focus of this paper will be on industrial property and on patents, more specifically. The paper will provide an overview of the major scholarly contributions that have contributed to define the properties of intellectual property institutions, with specific emphasis on public IP-related institutions. Section two will sketch out the broad economic rationale for public intervention for the provision of incentives to innovate. Section three will broaden the perspective by considering alternative public policy instruments for the provision of incentives to innovate. Section four will briefly review the main legal theories providing a rationale for the creation of an intellectual property system, while section five will further articulate the main elements of the economic approach to intellectual property. Section six concludes.
2. The rationale for providing incentives to innovative activity

The distinctive features of knowledge prevent it from being defined either as a pure private good or as a pure public good. In defining knowledge as a private good one can look at it as a commodity that possesses value both as a consumption good and as a capital good. However, as David (1993) among others, has pointed out, it is not a commonplace commodity, in that it is highly differentiated and has no obvious units of measurement. While defining the market value of knowledge as a pure consumption good does not pose particular problems since (think about the demand for various kinds of artistic creations), the same does not apply when knowledge is viewed as a capital good. In this case its value is given by the stream of benefits that it can yield in combination with other tangible and intangible assets and problems stemming from asymmetric and imperfect information may arise.

Knowledge displays also the two properties generally associated with the definition of a public good: it is non-rival and non-exclusive. Moreover, whereas there are sizeable costs in its original production, the costs associated with its reproduction and diffusion are low. The property of non-rivalness refers to the fact that the use of information made by someone does not diminish the potential or actual use of the same piece of information made by someone else. This is all the more evident if only we think at the instantaneous and widespread diffusion of information made possible by modern technologies. But recognition of this property is not a new insight. It is the point of departure of many of the early contributors to the debate over intellectual property rights. This characteristic of knowledge also constitutes one of the features distinguishing private property from intellectual property. The supply of any material good is finite and scarcity provides one of the possible justifications for private property. In fact, well-
defined rights over material objects respond to the need for the rational use of scarce resources such as land. In the absence of property rights both static and dynamic problems would arise. The static problems would derive from the overuse induced by the fact that users do not take into account the costs they impose on others and from a lack of access to economies of scale. The dynamic problems would stem from the absence of incentives to improve the resource if someone else might subsequently appropriate the land and benefit from the improvements without compensation. In the case of IPRs, given that intellectual objects are non-rival in consumption and that joint or repeated use of knowledge does not diminish its supply, the distinction between traditional property theory and intellectual property is easily perceived.

The second property of knowledge as a public good, namely its nonexclusive nature, also contributes to distinguishing intellectual property from traditional property. That knowledge is nonexclusive means that it may be difficult or impossible to prevent others from accessing or using information once it has been created. Once an invention, a formula or an idea is disclosed it is virtually impossible to avoid its unauthorized use in the absence of a legally enforceable claim over it. In the case of tangibles, it is possible for people to protect their property from violation by others. In the case of intangibles, it is only through law that possession can be guaranteed.

As for the third characteristic of knowledge, the combination of high fixed costs of production and low costs of diffusion suggests that there is a public interest in ensuring wide access to the information created. In fact, if the marginal economic cost of providing the information to an additional user approximates zero, social welfare calls for widespread diffusion.

Recognition of these properties of knowledge has historically led to widespread consensus concerning the idea that innovative activity takes place in a situation of market failure generated by the existence of an “appropriability problem”. Pigou (1920) is the first to expressly identify such problem and to link it to the public
good nature of information. Building on Pigou's analysis, Arrow's (1962) seminal article identified the existence of a situation of market failure by combining the appropriability problem first described by Pigou with the recognition that the marginal cost of increasing the utilization of information is zero. Arrow's analysis highlighted that the characteristics of knowledge drive a wedge between the social value of innovation and the value that is privately appropriable, thus generating an incentive problem that suggests the need for some form of public intervention.

Finally, it is worth considering two aspects that distinguish knowledge from conventional public goods such as defence and traffic lights. The first aspect relates to information asymmetries. Since the attributes of information cannot be known before information is transferred, it may be difficult to sign and enforce contracts concerning the production and diffusion of knowledge. The second aspect is given by the cumulative and interactive nature of knowledge. Inventions and intellectual creations constantly build upon previous work. Knowledge has a capacity for internal growth that has no equivalent in the case of classic public goods.

### 3. IP put in context: incentive mechanisms alternative to IP

One way for the State to fix the “appropriability problem” is to confer exclusive rights over the exploitation of newly created information for a limited time period, i.e. to grant intellectual property rights (IPRs). Intellectual property is information that possesses economic value mainly because, being made artificially excludable by law, it can be subjected to commercial exploitation. Conferring a right over intellectual property means granting the legally enforced power to make use of the information created, to exclude others from using it or to define the terms on which it can be used.
By increasing the extent to which the benefits of intellectual creations can be appropriated, IPRs are supposed to bring the private incentive to innovate closer to the socially optimal level. The instruments of intellectual property protection assume different forms according to the protected object. *Patents* protect new, non-obvious and commercially valuable inventions granting the patent-owner the legally enforceable power to make commercial use of them. *Copyrights* protect the original expression of an idea although they do not protect the idea itself. *Trademarks* protect words or symbols that identify for consumers the products of a particular firm. Similarly, *geographical indications* identify for consumers the products of a particular city or region. “*Sui generis*” forms of intellectual property protection include *Layout design of integrated circuits*, which protects producers of semiconductors, and *Plant Breeders’ rights*, destined to the protection of new plant varieties. *Trade secrets* are also generally considered as part of a system of intellectual property rights, in that they protect “any formula, pattern device or compilation of information which is used in one’s business, and which gives him an opportunity to obtain an advantage over competitors who do not know or use it”\(^1\).

Although by far the greatest amount of scholarly attention has been devoted to the analysis of the award of private rights to information as a mechanism to promote innovation, States have various instruments at their disposal to cope with the dynamic inefficiencies that may arise as a consequence of the public-good nature of innovative knowledge. Along with the intellectual property system, reward systems have also been adopted, both in the form of prizes and of procurement. The term “prize” refers to a publicly financed payment made to researchers who succeed in being the first to deliver a pre-defined invention and surrender any private right to it thus leaving it to the public domain. The term “procurement”, in turn, refers to the contracts the State stipulates to obtain

\(^1\) According to the American Law Institute’s definition.
specific innovations that may, or may not be destined to the public domain (David, 1993).

Reward systems have recently attracted renewed theoretical interest\(^2\). Shavell and van Ypersele (2001) and Gallini and Scotchmer (2002) provide an interesting comparison among alternative incentive forms, although the former authors do not distinguish between prizes and procurement and refer more generically to “reward systems”. The two papers converge with respect to some broad observations, namely that:

a. Intellectual property does not enjoy any fundamental advantage over reward systems, however defined.

b. Reward systems possess the advantage of avoiding the monopoly welfare loss associated to intellectual property rights.

c. The incentive provided by intellectual property is more closely linked to actual social surplus than that provided by rewards, as firms take the decision to invest by comparing some measure of the value of innovation to its cost.

d. The principal difficulty associated to the adoption of a reward system resides in the sponsor’s need for information to select the most efficient firms and to calculate the appropriate magnitude of the reward.

Shavell and van Ypersele are more sanguine about the opportunity of adopting some form of reward system. In particular, by constructing a model in which a single potential innovator possesses ex ante knowledge about the demand curve for the innovation, whereas the government knows only the probability

\(^2\) Rewards have attracted scant attention in the course of the past century. Two interesting contributions on the subject are Polanyi (1943) and Wright (1983), both making a case for preferring rewards over patents provided the State can gather sufficient information. More specifically, Wright constructs a formal argument showing that the relative desirability of patents, prizes and procurement rests on the nature of the private information. When the informational asymmetry concerns costs, prizes are superior to patents in that they allow to achieve the same results but save on the monopoly deadweight loss, whereas if the informational asymmetry concerns the value of the innovation patents may offer an advantage because the incentive they provide is linked to actual social surplus. Arrow (1971) also considers some merits of rewards. Nalebuff and Stiglitz (1983) evaluate the role of prizes, concluding that they may provide strong incentives for research but tend to generate a duplication of research efforts.
distribution of demand curves, they suggest that an optional reward system is superior to the patent system because it allows to save on the monopoly deadweight loss and may address the problem of overinvestment in research by an appropriate selection of the magnitude of the reward^3.

The more nuanced analysis by Gallini and Scotchmer reveals some important distinctions between prizes and procurement and considers the issue of the choice of the most efficient incentive mechanism when there is more than one potential innovator. The shift of focus from a single potential innovator to multiple potential innovators puts emphasis on the problem of aggregating information about the value and cost of potential innovations. To this respect, neither IP nor prizes or simple procurement mechanisms are effective. They distinguish between the decision problem, i.e. the choice of whether a given innovation should be pursued, and the delegation problem, namely the issue of the choice of the most efficient innovator(s) and of the efficient rate of investment. They conclude that a) the relative performance of "simple” mechanisms varies according the characteristics of the environment considered, and particularly according to the availability of information on innovations’ costs and values; b) a form of procurement that allows to provide payments restricted only to certain firms or differentiated according to firms’ relative efficiencies is preferable to IP or simple prizes.

In addition to the simple incentive mechanisms described above, proposals for the implementation of more creative mechanisms have also been advanced. Among them, at least the following are worth of consideration. Gandal and Scotchmer (1993) propose that the sponsor offer a menu of options with both fixed fees and firm-specific prizes.

Kremer (1998) considers the possibility of implementing a mechanism based on patent buyouts, especially in the pharmaceutical industry. The problem of

---

^3 In addition to this, they also show that a simple reward system may be dominated by the patent system, whereas in circumstances in which the simple reward system dominates patents, it might also be superior to the optional reward system.
determining a price appropriately reflecting the invention’s value is addressed through an auction. The price thus obtained, times a markup that would reflect the difference between the social and private value of invention, would be offered by the government in return for placing the patent in the public domain. However, in order to induce truthful revelation of firms’ evaluation, the government would also sell to the highest bidder a small proportion of patents, randomly chosen. One important cautionary note regarding Kremer’s proposal is suggested by Gallini and Scotchmer (2002), who observe that such a scheme would be efficient only if there is a single researcher, as the presence of multiple potential innovators would involve the possibility of excessive and/or inefficient investment.

In line with Kremer's analysis is also the recent analysis by Pagano and Rossi (2009) who also propose a mechanism of patent buyouts but focus on the economics implications of the adoption of such a policy tool at a time of economic crisis rather than focusing on the details of its implementation. The two authors highlight the possibility to reconcile short-term and long-term policy interests by activating a "super-multiplier" effect connected to the release in the public domain of the information embodied in valuable patents and explore advantages and disadvantages of such a mechanism.

Hopenhayn Llobet and Mitchell (2006) examine an innovative environment characterized by cumulative innovation and heterogeneous research ability. Assuming incomplete information on the part of the patent office, which results in the possibility of both moral hazard and adverse selection, like Kremer they propose an incentive mechanism that combines patents and buyouts. Differently from Kremer, however, firms are offered a menu of patents with infinite duration and fixed scope, so that bigger improvements get greater protection. To implement this mechanism they suggest a system of buyouts or licensing fees in which innovators commit at the outset to a price at which they will relinquish their rights. Such a mechanism should allow the patent office to obtain information concerning patent value and would only require information about whether an
improvement has occurred and not about the amount of the improvement to be implemented.

Compulsory licensing is also part of the mechanism suggested by Kingston (2000). Kinston’s creative proposal for reform includes:

- The granting of an “innovation warrant” consisting in an exclusive privilege irrevocable for a short period of time and granted to any intellectual creation responding to the lax novelty criterion of being “not available in the ordinary course of trade”.
- A term of privilege variable according to the risk of the innovative investment.
- A system of compulsory licensing characterized by payments that are a) a socially acceptable multiple of the R&D investment the initial innovator has made, and b) a one-off capital sum rather than royalties.

What is perplexing about Kinston’s scheme is, among other practical aspects of its implementation, the burden of administrative costs it would impose on regulatory agencies. In particular, in spite of Kingston’s optimist attitude towards the government’s ability to define “a socially optimal multiple” of the R&D investment as a basis for determining the compulsory licensing fee, one is left wondering whether the problem of asymmetric information that is prevalent in the literature on rewards can be dismissed so easily. Additional doubts arise because a) R&D costs do not always reflect the quality of the invention, so that the payment Kingston envisages would be different across industries and sectors but would not be differentiated according to firms’ relative efficiencies and innovations’ social value; and b) a payment based only on R&D costs might encourage firms to artificially inflate them.

This brief review of the mechanisms that have been proposed as an alternative to the patent system in its current form suggests that it has probably raised more questions than it has answered. An interesting avenue for future research would be to explore the practical implications of the few conclusive answers available by appropriately linking them to specific industries and sectors. Additional research
is needed to redress the balance of the theoretical interest on incentives to innovation towards mechanisms alternative and/or complementary to intellectual property protection.

4. Intellectual property rights: the legal perspective

The legal theoretical debate is animated by the interplay among four approaches: Utilitarianism; Labor Theory; Personality Theory and Social Planning Theory. The reason of their prominence largely depends on the fact that their building blocks are constituted by concepts and lines of reasoning adopted by judges and lawmakers in current law practice. This is especially true for Utilitarianism in the context of U.S. law and for Personality Theory in European (but most of all French and German) law. However, as Fisher (2001) has pointed out, “theorists are seeing the law through glasses supplied by political philosophy”. This means, on one end, that the four approaches are seen as irreconcilable by theorists, whereas the concepts underlying each of them are all present and generally blended in legislative and judicial materials such as constitutional provisions, judicial opinions, case reports etc. On the other end, some of the issues that occupy a relevant space in intellectual property law and practice do not find comparable emphasis in legal theory. The reason of this distance between intellectual property law and legal theory is that legal theorists apply to debates over intellectual property the same categories that shape their discussions over political philosophy or property law in general.

4.1. Utilitarianism

The utilitarian approach is the one that has gained the strongest support among legal theorists as much as among economists. It is rooted on the works of authors such as Jeremy Bentham, John Stuart Mill and A.C. Pigou. In line with
the Benthamite ideal of the “greatest good of the greatest number” utilitarianism poses as the objective of a system of property rights the maximization of social welfare. In the case of intellectual property such a goal requires the definition of rights over intellectual objects so as to provide incentives to the producers. The main line of reasoning proceeds from the assumption that intellectual creations present such distinctive features – e.g. non-excludability, low costs of reproduction - that make difficult for the inventor to obtain compensation for his inventive effort. In the absence of intellectual property rights - the argument continues- it would be impossible for creators to appropriate the benefits arising from their work and to recover the costs sustained. As a consequence, they would have no incentive to invent in the first place. In maximizing social welfare, however, it must be taken into account that granting exclusive rights over an invention to its creator limits the access to the invention by others. Designing an appropriate system of intellectual property rights requires thus to strike an optimal balance between the need for providing incentives to innovations and the drawback of reducing their public enjoyment.

The utilitarian approach is teleologically oriented, the objective being the enhancement of social welfare through technological progress. This statement requires a couple of remarks. The first aspect that it is worth noting is that it should not be taken for granted that the target assumed is invariably a “good” for society. The basic assumption of the utilitarian perspective is that the short-term drawback of granting monopoly rights to an inventor has to be accepted in order to achieve the long-term development of society. This, of course, raises the question whether long-term development is dependent on the ever-accelerating pace of technological progress. There are opportunity costs to investing in scientific research and inventive activity. The empirical research conducted so far has not provided definitive answers neither on whether the stimulus to innovation provided through the granting of intellectual property rights is worth its cost nor on whether intellectual property rights allow the minimization of the social costs entailed by innovation. Secondly, it must be taken into account that the incentive
effect of intellectual property rights operates at different degrees in different fields of technology. The extent to which technological progress contributes to the long-term development of society at large crucially depends on which of the technological fields will be affected the most. The commitment of utilitarianism to technological progress does not account for the difference for society’s development between the invention of a life-saving drug and the creation of a new kind of musical equipment. Finally, utilitarianism deserves a warning because of its nature of “ends-justifies-the-means” philosophy. History clearly shows the shortcomings of such a philosophy and imposes deep reflection before accepting it.

Even if we disregard the problems posed by the teleological orientation of utilitarianism and we accept in toto the end of enhancing welfare through technological progress, we face an additional problem deriving from the limited prescriptive power of this approach. What criterion should be adopted to choose among different legal orderings? What is the rule that allows translating into more concrete terms the objective of welfare maximization? Among the many criteria proposed two stand out because they are chosen by a majority of contemporary writers. The first is the “wealth-maximization” criterion that adopts a measure of welfare based on consumers’ ability and willingness to pay for goods and select the system of rules that maximizes aggregate welfare. The second is the “Kaldor-Hicks” criterion. A set of legal rules is Kaldor-Hicks superior to another if the gains it generates are such that, by moving from the second set to the first, all those made worse-off by the policy can be compensated through a lump-sum transfer and some are still made better off.

Neither of these criteria can be accepted without objections. A common objection to the two criteria is that the narrow definition of social welfare that underlies them needs to be questioned and deepened. Unresolved theoretical problems related to the adoption of these two criteria are also those arising from the measurement of utility, the need for interpersonal comparisons of utility and the aggregation of utility functions. These issues lie at the heart of social choice
theory and are not at all limited to the field of intellectual property rights. Consequently, if it is useful to recall the existence of controversies over the theoretical foundations of the two cited criteria, a deep analysis of the treatment that social choice literature reserved to this topic is not an aim of this work. However, I find it useful in this context to recall the central criticism moved to the net utility calculus that is the core of the utilitarian approach. A policy based on a net utility calculus does not take into account differences among individuals. As Amartya Sen (1985) has argued, the consequence is that “the elimination of society’s ills is justified in utility theory only if there is a net utility through their removal”.

4.2. Labor Theory

Labor Theory justification for intellectual property rights has its origin in a Lockean-style natural law argument. The creator of an intellectual object has a natural right upon it because of the effort he spent to put it into being. In other words, the mixing of labor to resources “held in common” naturally leads, as a matter of fairness, to attribute a property right to the laborer. The attribution of property rights encounters, in Locke’s theory, only two limitations (the so-called provisos). The first limitation requires that the others are not made worse-off by the attribution of property rights. The second proviso poses limits to the amount of an object that can be used by a person, setting the limit at the point before it spoils.

This argument, originally elaborated for tangible property, is seen as particularly suited for intellectual property. The reason is twofold. First, labor’s contribution to the value of the intellectual object is greater than the raw materials’ contribution. Second, the raw materials used for intellectual creations are things such as ideas and concepts, for which the property of being held in common appears evident. The extension of the Lockean rationale to intellectual
property is not however as straightforward as it may first appear. On one side, the interpretation of Locke’s theory is not free from ambiguities so that it fails to provide univocal guidelines for the shaping of an intellectual property system. On the other side, the system of property rights that would result from the application of the few unambiguous prescriptions that can be derived from Locke’s scripts is substantially different from the existing one. The latter cannot therefore be justified on the grounds of Locke’s theory.

As for the first aspect of the problem, many questions fail to find a definite answer. First of all, why exactly does the use of labor upon a common resource justify the attribution of a right over the resource itself? William Fisher lists a great number of reasons that can be found scattered in the Second Treatise. Not all of them provide to the same extent a justification for intellectual property. More specifically, what is the main line of reasoning used to justify tangible property does not apply equally well to intellectual property. The Lockean argument that envisages individual appropriation of resources as a consequence of the natural right that man have to their preservation is based on the assumption that resources are scarce. This assumption, as noted before, does not hold perfectly in the case of intellectual resources, because of the properties of non-exclusiveness and non-rivalness in consumption. Suppose we can dismiss this first problem. Other problems would arise from the interpretation of Locke’s theory. Locke has provided neither a definition of “intellectual labor” nor one of the “raw materials” used to create intellectual objects. The features of a system of intellectual property rights based on Locke’s theory depend significantly on the interpretation we give to these concepts. The fact that they haven’t been clearly specified leads to a great extent of indeterminacy, which limits the prescriptive power of Locke’s approach to intellectual property.

The second problem that arises when trying to justify the system of intellectual property rights on the basis of Lockean arguments is that a system of intellectual property rights compatible with the few straightforward prescription that can be deduced by the Treatise would be quite different from the existing one.
First of all, the creator of an intellectual object acquires the property of the new invention he made through his work, but he does not appropriate the resources originally held in common, i.e. the existing concepts and ideas he used for the invention. Secondly, intellectual property lacks the properties of exclusivity and eternity that Locke envisaged for tangible property. A right upon an intellectual creation is not as exclusive as a right upon a piece of land, in that it prevents others from copying the idea but not from utilizing it in order to create a different intellectual object. Moreover, the current system of intellectual property protection grants rights whose duration is limited, in contrast with the natural outcome of Locke’s arguments.

One additional problem is posed by the definition of ownership in the cumulative inventive process. Given that inventions are usually built one upon the other, and that each invention uses materials that are the fruit of the labor of others, how can the limits of ownership of each intellectual object traced? Should a partial right to the new invention be granted to the inventor of any intellectual creation used in its elaboration? Locke’s theory does not provide any conclusive answer to this issue.

The difficulties posed by the indeterminacy of Locke’s justification of intellectual property have fostered a lively debate among the writers that place themselves in the school of thought of Lockean ascendance. One of the most important exponents of this theoretical orientation is Robert Nozick. He switches the emphasis away from labor and towards Locke’s first proviso, i.e. the assumption that a person may be granted an exclusive right upon the fruit of his labor as long as “there is enough and as good left in common for others” (Locke, 1972). The acquisition of property through labor is thus justified if and only if others are not made worse-off by the appropriation, the baseline for the comparison being the absolute condition of people before the appropriation. Nozick’s interpretation of the condition of being “worse-off” is of a weak type. He includes in the definition losses such as a decrease in one’s own wealth or a restriction of the set of resources that one can freely use, but does not take into
account the harm constituted by the lost opportunity to acquire property rights upon previously unknown resources.

An interpretation of the first proviso of this sort is compatible with the patent system. To see this we can refer to an example that Nozick makes in *Anarchy, State and Utopia*. He makes the case of the issuance of a patent for a new useful drug. The researcher that created the new drug is entitled to an exclusive right over its creation because, even if a patent limits the ex-post accession to invention by other people, in the absence of a patent the invention would not have been created. Moreover, the restriction of the set of resources available for use is of limited importance, because the substances used by the researcher for his discovery are likely to be abundant in nature. As a consequence, the patent will make at least some people better off with respect to the situation prior to the discovery. More specifically, the people that would be made better off are the ones that will be able to buy the drug and use it to recover from their disease.

This last consideration naturally leads to question the interpretation of being left “worse-off” adopted by Nozick. As noted, Nozick’s analysis reads the condition of being made worse-off in an absolute sense, the baseline of comparison being the absolute condition of people before the establishment of property rights. Ostergard (1999) stresses the point that the “worse-off” condition must be interpreted in a relative sense when using it to choose among different legal orderings, because it is with the collective good that we must be concerned in similar contexts. In other words, the effects of the attribution of property rights must be analysed taking into account the situation of people after the appropriation relative to the situation of others (*under the same circumstances*). In the example of the researcher, the fact that granting intellectual property rights upon the discovery implies allowing the researcher to charge a monopoly price limits the accessibility to the drug only to those who can afford it. As a consequence, a fraction of the population that would benefit from the drug is “worse-off” relative to the people that have enough financial resources to
purchase the drug. If there remarks are taken into account, Nozick’s theory displays relevant limits in justifying intellectual property rights.

4.3. **Personality Theory**

The personhood approach to property rights has its roots in Kant’s Philosophy of Law and Hegel’s Philosophy of Right. The basic idea of this line of reasoning is that the establishment of private property is functional to the protection of fundamental human needs and interests. In the words of one of the most representative contemporary contributors to this theory: “the premise underlying the personhood perspective is that to achieve proper development – to be a person – an individual needs some control over resources in the external environment. The necessary assurances of control take the form of property rights.” (Radin, 1982. The arguments in favour of intellectual property rights stemming from this premise are wide-ranging. One option is to look at intellectual creations as means of expression of authors’ and inventors’ “wills”. Another is to emphasize the role played by intellectual property rights in fostering creative intellectual activity and human flourishing through it. Whatever is the specific aspect of this theory that is emphasized, there is at least one guideline for the shaping of an intellectual property system that is common to all the approaches. The more an intellectual object reflects its creator’s personal identity the more it deserves legal protection. Following this assumption, a song merits protection more than the formula of a drug, for example, because the former is considered to express the author’s personality more significantly than the latter.

Despite the fact that they share this common assumption, proponents of the personhood approach take different views with regards to many specific aspects of the design of an intellectual property regime. The controversies are wide and varied and it is sufficient to recall just two topics to perceive their depth. There is
no agreement neither on whether an inventor can alienate his right to control the copying of his work, nor on whether an author is allowed to prevent others from imitating his work. The existence of such controversies is the natural outcome of the major limits of this approach. The limited prescriptive power of this perspective depends on two interrelated problems. The first is that even though it may be plausible to design a system of intellectual property rights to protect fundamental human needs, it is not at all evident what interests exactly deserve protection. The actual shape of a system of property rights will depend substantially on what needs are considered to be crucial. Related to this is the fact that the concept of personhood that lies behind this approach has not been extensively and concretely defined. The result is that any effort to derive concrete guidance for the shaping of the intellectual property rights system suffers from a high degree of abstractness.

4.4. Social Planning Theory

The objective that proponents of this approach choose as the basis for a system of property rights is the achievement of a just and attractive culture. What has to be meant for just and attractive culture is not a matter of secondary importance. It is a question at the heart of the theory and it is at the very heart of the theory that its limitations appear evident. In defining the concept of culture that has to be cultivated through the property system exponents of this theory refer to various ascendants from the political and the legal field. Among those there are Jefferson, the early Marx and the Legal Realists. Whatever is the philosophical stance taken, the basic problem lies in the arbitrariness of the choice. In a recent essay considering intellectual property issues on the Internet W. Fisher, advocate of the Social Planning Theory, sketched out his vision of an attractive intellectual culture. The picture included aspects such as consumer welfare, distributive justice, and access to a broad range of intellectual products and participation to the
cultural process. Referring to concepts such as the ones listed above inevitably leads to face controversies that animated political and philosophical debates for centuries. Since these controversies are bounded to remain entwined with the main arguments provided by Social Planning Theory in favour of IPRs, the inadequacy of the justification that this theory adduces for IPRs results evident. The unavoidable arbitrariness that lies at the heart of Social Planning Theory explains also why this approach has frequently had to respond to the accusation of being “illiberal” and “paternalistic”.

5. *Intellectual property rights: the economic perspective*

5.1. *The economic functions of the patent system*

5.1.1. IP incentive function

As mentioned, the primary rationale for granting patent protection is to provide incentives to innovate to economic actors facing difficulties in appropriating the returns from their intellectual creations - the so-called incentive function of patents. Society is ready to grant a limited monopoly on the newly created ideas in return for the inventor's innovative effort, so that a trade-off between monopoly costs and benefits in terms of increased innovation is established. The convenience of incurring such trade-off, however, has long been put into question by those emphasizing the virtues of competition as a stimulus to innovation. This alternative view, expressed among others by Kenneth Arrow (Arrow, 1962) and more recently by Boldrin and Levine (Boldrin and Levine, 2008), stresses the existence of incentives to innovate other than state-created monopolies over ideas and suggests that monopoly may be less effective than competition in creating an environment conducive to innovation.
The possibility that both of these views have some merit cannot be ruled out altogether. Empirical studies of the mechanisms employed to appropriate innovation returns have shown that patents play a very different role in different industries (Mansfield, 1986; Taylor and Silberston, 1973; Levin et al., 1987; Cohen et al., 2000) indicate, for instance, that the importance attributed to patents is high primarily in pharmaceuticals, while a combination of legal and non-legal mechanisms such as lead time, secrecy and possession of complementary capabilities plays a more decisive role in all of the other sectors considered.

The effectiveness of the incentive function will, in general, depend on the underlying characteristics of innovative knowledge. Recent economic scholarship has recognized that, in contexts characterized by cumulative and complementary innovation, patents may create some additional problems with respect to industries producing products of a more discrete nature. Cumulativeness or sequentiality refers to the fact that innovations build upon each other and each innovation constitutes an incremental improvement over the previous one. Complementarity indicates, at the same time, the existence of some sort of positive externality among the research paths taken by different innovators, i.e. by taking a different research line each potential innovator increases the overall probability of success in a given innovative endeavour, and the fact that often a successful innovation requires the combination of multiple complementary inputs.

In order for patents to really serve their purpose of promoting innovation, patent design needs to optimally solve the problem of dividing profits between pioneers and improvers, and between successive innovators in general (Scotchmer, 1991; Chang, Green and Scotchmer, Scotchmer, 1999). Increasing protection for early innovators raises the costs of improvements. If an appropriate balance is not found, it is possible for patents to constitute a drag, rather than a stimulus, to innovation.
In a study concerning the software industry, for instance, Bessen and Maskin (2000) conclude for the greater effectiveness, under circumstances of sequentiality and complementarity, of a setting characterized by competition and rapid imitation on the basis of both a dynamical model and some empirical observations concerning the natural experiment occurred in the U.S. software industry as a consequence of the courts' change of attitude towards software patentability in the 80s.

5.1.2. IP transactional function

Patents aim to solve the well-known Arrovian paradox, allowing for the exchange of intangibles. This is true not only when intangibles are in the form of end-products, but especially when they constitute inputs into further activities. Thus, the existence of patent protection over the inputs to a collaborative research endeavor may facilitate inter-firm R&D collaboration. Small firms that lack resources to effectively commercialize their inventions may access the market by becoming specialized technology suppliers. By assigning ownership entitlements to successive innovators, patents determine bargaining positions that facilitate welfare-enhancing transactions and enable some sort of division of profits among subsequent contributors to a given stream of research⁴. A more radical view holds that, by enabling bargaining, the assignment of strong and broad patent rights may provide incentives for further product development and refinement and may help to coordinate development effort in such a way so as to reduce the duplication of innovative efforts. This is what Edmund Kitch (1977) has dubbed the "prospect theory" of patents.

What proponents of the above view emphasize as an advantage of patents, namely that patents make licensing indispensable in order to combine existing technologies into new ones, represents a cause of concern for other scholars. In

---

⁴ This is the rationale underlying, for instance, the U.S. doctrine of blocking patents, intended as a means of "forcing" transactions.
the context of biotechnology, for instance, Michael Heller and Rebecca Eisenberg have warned about the potentially disastrous consequences of the excessive fragmentation of intellectual property entitlements - the so-called "anticommons" problem (Heller and Eisenberg, 1998). When development depends on the combination of multiple components each of which may be the object of patent protection, as it is the case for complex and cumulative technologies like biotechnology and software, transactional hurdles may end up exerting an excessive "tax" on future innovations.

As far as transactional issues are concerned, "anticommons" represent only one side of the coin, the other being the creation of patent portfolios and the adoption of "patent thicket" strategies. The mentioned Carnegie Mellon survey (Cohen et al., 2000) suggests that firms involved in innovation in "complex" technologies tend to accumulate portfolios of patents to be used as bargaining chips in negotiation. The accumulation of patent portfolios may, in some industries, be the only way around the problems posed by the existence of "patent thickets" and at the same time may contribute to enhance such problems. Difficulties associated to the existence of "patent thickets" arise when the rights over components necessary to develop a marketable invention partially overlap, so that no developer is free to commercialize his invention without infringing, or potentially infringing, someone else's patent. Building a portfolio of patents on related technologies can both protect firms from potential hold-ups and provide a means of limiting entry by competitors. The problem is related, but not identical to the "anticommons" problem.5

Under such circumstances, patent pools and patterns of cross-licensing generally develop to enable continued innovation. This has been proved, for instance, for the semiconductor industry, where Hall and Ziedonis (2001) have found the increase in patenting to be strongly related to the adoption of patent portfolio strategies and have highlighted the existence of clear patterns of cross-licensing

---

5 For a clear distinction between the two phenomena see Burk and Lemley (2003).
among established industry players. Evidence of the strategic use of patents as bargaining chips and of the increase in patenting associated to these patenting strategies exists also for the IT field. One study relevant in this regard is the OECD/BIAC survey. This study finds, among other things, that more than three-quarters of the ICT firms surveyed now patent technologies they would not have patented one decade ago, even if patentability was then unrestricted (Sheelan, Guellec and Martinez, 2003).

5.1.3. IP disclosure function

The patent system is meant to perform a disclosure function that is often neglected in theoretical analyses but enjoys a relatively high degree of popularity with courts. In exchange for a range of exclusive rights over inventions, patent-holders are required by statute to disclose their protected inventions to the public so as to allow an effective diffusion of technological knowledge. This exchange is often referred to as a bargain between inventors and the State and its rationale is generally referred to as the “contract theory” of patents (see, e.g., Denicolò and Franzoni, 2004). According to this theory, patent protection increases the availability of scientific and technological knowledge that would otherwise be kept secret, inducing both direct benefits in the form of increased knowledge diffusion and indirect benefits in the form of a reduction of wasteful duplication of innovative efforts.

The general principle of patent disclosure is declined differently in different legal contexts, although its essence tends to be the same in any patent system. Art. 29(1) of the TRIPs Agreement provides that “[m]embers shall require that an applicant for a patent shall disclose the invention in a manner sufficiently clear and complete for the invention to be carried out by a person skilled in the art…”. Provisions similar to the one set forth in the TRIPs Agreement appear, for
instance, in art. 83 of the European Patent Convention⁶ and in 35 U.S.C. § 112, para. 1⁷. At the core of these provisions there is a common criterion of “enablement”, requiring inventors to enable anyone skilled in the art to practice and reproduce the invention, which has become a worldwide minimum standard of adequacy of disclosure (Reichman, 1995). This criterion serves at the same time the purpose of ensuring the accessibility to the public of information concerning the invention and the purpose of avoiding that the patent monopoly be granted in exchange for a contribution to technological progress effectively achieved by the inventor. This suggests the existence of some correlation between the scope of disclosure and the scope of claims, although it seems that a consensus is emerging on both sides of the Atlantic on the fact that this correspondence should not be considered excessively strict (Janis, 2000).

The exact content of the disclosure requirement is difficult to spell out. Multiple doctrines have developed in different jurisdictions in order to clarify the implications of the disclosure requirement, especially with regard to its application in technological domains recently admitted to patentability, such as biotechnology (Kopczynski, 2002)⁸. The development of these doctrines has not led, however, to the emergence of an agreed-upon standard for disclosure.

Patents disclosure function is surrounded by some degree of controversy also from an economic standpoint. The notion that the patent system plays a positive role in inducing disclosure has been criticized on various grounds. Of course, the greatest criticism relates to the fact that patents are not meant to be a means of information dissemination: they are written by patent attorneys motivated by the objective to obtain exclusive rights as broad as possible and with no incentive to

---

⁶ Art.83 EPC provides that “[t]he European patent application must disclose the invention in a manner sufficiently clear and complete to be carried out by a person skilled in the art”.

⁷ 35 U.S.C. § 112, para. 1 provides that “[t]he specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor for carrying out his invention” (emphasis added).

⁸ We refer here, among others, to the “written description requirement” in the United States and to the notion of “support” embodied in art.84 of the European Patent Convention.
disclose information useful to the firm’s competitors. A number of more specific issues have been also raised. On one side, the concrete availability of the option to keep the invention secret has been questioned (Plant, 1934). On the other side, the argument that only inventions that cannot be kept secret are in fact patented has been historically advanced (Machlup and Penrose, 1950; Bessen, 2005). Moreover, some theoretical models of patenting behaviour have also highlighted the possibility that the disclosure obligation might induce inventors not to patent (Horstmann et al., 1985; Scotchmer and Green, 1990). More recently, some new insights on the social value of patent disclosures have been offered by authors emphasizing the effects of disclosure on rent-seeking behavior and as means to convey information on new uses of a given technology (Landes and Posner, 2003).

The effectiveness of patents disclosure function is, ultimately, an empirical question. The answers to this question available to date, tend to confirm the view that, in spite of the availability of patent protection, secrecy constitutes a relevant mechanism of appropriation of the value of innovations, one that is actually much more relevant than patents in most industries, especially in the protection of process innovations (Cohen et al., 2000). Moreover, a comparison of survey results gathered in 1994 by Cohen and others (Cohen et al., 2000) with a similar survey conducted in the early 1980s (Levin et al., 1987), shows that recourse to secrecy in the protection of product innovations has increased over time across a range of industries.

The question of the effectiveness of patents disclosure function has also been addressed from a different perspective, by studies aimed at exploring whether information disclosed through patents and available in patent databases are of any use to firms in different sectors. In principle, patent-induced information might be useful to firms in various ways: as a means to monitor technological advances in their own sector; as a way to identify new applications of existing technologies in fields unrelated to the one in which they were developed; and as a way of gathering relevant legal information, such as information on the likelihood that
one’s own patent infringes someone else’s patent or vice-versa. In practice, patent databases are rarely consulted for reasons other than legal purposes (Oppenheim, 1998; Puay et al., 2001). This holds particularly for small and medium enterprises (SMEs). A survey of SMEs conducted in Europe by Arundel and Steinmuller (1998) shows, for instance, that patents are a relatively unimportant source of technical information for the surveyed firms, although its relevance increases with firms’ size, sector of activity and propensity to patent. The limited relevance of patent-induced information disclosure for SMEs depends, in most sectors, on the high costs involved by the expert consultation of patent databases. In the IT sector, to this reason should also be added the perception that the 18 months delay before publication of the patent application makes published information relatively obsolete when made publicly available.

5.1.4. IP signalling function

In addition to the previously mentioned functions, patents have also been credited with a "signaling" function. Possession of patents, according to this view, serves the purpose of signaling a firm’s innovative capabilities and increases its ability to raise the necessary capital, especially through venture capital financing. This function is, of course, critical to SMEs, who may not have much opportunities alternative to venture capital investors to find the capital necessary to commercialize an invention and enter a new market. The link between patenting and venture capital financing has been highlighted as fundamental especially in biotechnology.9

9 See, for instance, Coriat and Orsi (2002), arguing for the existence of an "institutional complementarity" between the U.S. intellectual property system and the nature of financing - venture capital financing - prevailing in the most innovative milieu of the U.S. economy.
5.2. The optimal design of the patent system

A consistent body of economic literature has explored the characteristics of the optimal design of the patent system, starting from analytical settings envisaging length as the relevant policy instrument and a single inventor operating in isolation and subsequently extending the analysis to take account of a broader range of policy instruments, multiple innovators and the possibility of cumulative innovation.

The classic treatment of the problem of optimal patent duration is provided by Nordhaus (1969) and Scherer (1972). Both authors determined in a static and deterministic framework and referring to cost-reducing process inventions the optimal patent length. They showed that the optimal duration of a patent strikes a balance between the size of the incentive effect of innovation and the dimension of the deadweight loss due to monopoly. One relevant and often neglected result of Nordhaus’ model is that the optimal patent life is extremely sensible to changes in the parameters of the system, and especially to the price elasticity of demand in the end-product market and to the efficiency of the R&D investment. As a consequence, the optimal duration of a patent is not uniform across industries or across inventions.
Nordhaus’ seminal paper on optimal patent duration

The classic treatment of the problem of optimal patent duration is that elaborated by Nordhaus. Nordhaus determined in a static framework and referring to cost-reducing process inventions the optimal patent length. He showed in a formal model that the optimal duration of a patent is the finite number of years that determines a balance between the size of the incentive effect of innovation and the dimension of the deadweight loss due to monopoly. The main result of Nordhaus’ model is that the optimal patent life is extremely sensible to changes in the parameters of the system. As a consequence, the optimal duration of a patent is not uniform across industries or across inventions. More specifically:

1. The higher the price elasticity of demand, the shorter the optimal length of the patent. This is because, at least for a certain number of years, the deadweight loss is an increasing function of the elasticity (the higher the elasticity, the greater the reduction of demand and consequently the deadweight loss)

2. The optimal duration of a patent is shorter if the invention is easy, i.e. if the cost-reduction obtainable for a given level of investment in R&D is substantial

3. The higher the social benefit stemming from the invention with respect to the amount of R&D effort, the shorter is the optimal patent life.

In his seminal paper Nordhaus also stressed an important point that should be borne in mind in considering present-day policy options concerning the patent system. Nordhaus stressed that, for inventions of limited importance or markets with low elasticity of demand, the patent system, while being a second-best solution, has a reasonable level of efficiency. However, he also highlighted that the reverse side of this statement is that the patent system results inappropriate exactly in the circumstances in which there is a higher potential for welfare losses. In light of this, it might be useful to whoever is interested in ascertaining the desirability of a patent system, to recall the words of Nordhaus: “the determination of the optimal patent life is ... not necessarily very important. What may be more important is to determine when the patent system has so low a level of efficiency that other mechanisms for encouraging technological change should be substituted for it.”

Nordhaus’ and Sherer’s analyses focus on the case of a single inventor and assume away the existence of R&D competition. Kamien and Schwartz (1974) and DeBrock (1985) consider the issue of optimal patent breadth in a setting
characterized by R&D competition. However, whereas in Kamien and Schwartz’s model the introduction of R&D competition suggests the need for a longer patent duration, as competition has the effect of reducing the expected payoff to each competitor, in DeBrock’s model competition reduces optimal patent duration. This is because DeBrock assumes that competition takes place at the research rather than at the development level, so that a longer patent life would induce unnecessary duplication of research efforts.

Patent scope is a policy variable that has attracted a growing amount of attention in the course of the past decade. O’Donoghue et al. (1998) rightly distinguish two dimensions of patent scope: breadth (also called “lagging breadth”) and height (also called “leading breadth”). The first refers to the protection the patent grants with respect to potential imitators and identifies the range of products that fall into the patent claim. The second refers to the protection the patent-holder enjoys with respect to subsequent innovations and improvements and is also referred to as the “novelty standard” or the “leading breadth” of a patent. Patent scope is subject to the discretion of both Patent Offices – that influence the extent of the claims an inventor is allowed to obtain at the moment of patent grant – and the judiciary – that provides an ex-post interpretation of infringement and is thus an instrument that can immediately reflect changes in policy, although probably with limited accurateness.

The first formal treatments of the issue of optimal patent breadth - Gilbert and Shapiro (1990) and Klemperer (1990) – are characterized by the adoption of a static framework and are much in the spirit of Nordhaus’ analysis of optimal patent length. Gilbert and Shapiro (1990) make the point that optimal policy calls for infinitely-lived but narrow patents – adjusted so that the patentee makes at least zero profits – whenever patent breadth is increasingly costly in terms of deadweight loss. Klemperer (1990) reaches opposite conclusions in a model of horizontal product differentiation. He demonstrates that broad and short patents have to be preferred to wide and long-lived ones if the width of the patent, allowing a price increase, induces consumers to switch to less preferred varieties.
of the products or to stop completely the consumption of the product. The definition of patent scope is different in the two models. Gilbert and Shapiro adopt as a proxy the flow rate of profits available to the patentee. By contrast, in Klemperer’s model the proxy for patent breadth is the region of the product space covered by the patent grant.

A case for broad and short patents is also made by Gallini (1992) in a static model that introduces the issue of wasteful imitation. In her model longer patent life encourages imitation, whereas increased patent breadth – defined as the cost that rivals must incur to imitate the invention avoiding infringement – tends to prevent it. The optimal combination of length and breadth results from the double objective of providing sufficient incentives to innovate and of deterring imitation.\(^\text{10}\)

The most apparent limit of the analyses of patent breadth reviewed so far is that they do not take into account the incentives of follow-on innovators, that is they focus on patented innovations as output of R&D but not as inputs for further research. Other models have incorporated this aspect into the analysis, thus addressing the issue of patent “height” in a cumulative context. Cumulative innovation renders relevant to the determination of optimal patent scope two additional objectives. The first is to facilitate timely disclosure of innovations, given that early innovators create a positive informational externality that benefits subsequent ones. The second is to balance the innovation incentives of “pioneers” and “improvers”, appropriately dividing profits among them, as an increase in the extent of protection granted to the former reduces the incentives of the latter by making it more costly to build on previous inventions.

As for the first aspect, Scotchmer and Green (1990) argue in favour of a weak novelty requirement because it encourages disclosure of even the smallest innovations. However, in their model firms are allowed a choice between

\(^{10}\) Other analyses of patent breadth in a static setting include: Denicolò (1996), Hopenhayn and Mitchell (1998), Maurer and Scotchmer (2002).
patenting and trade secrecy, and a weak novelty requirement encourages recourse to the latter, as disclosure benefits competitors. Matutes, Regibeau and Rockett (1996) also consider the problem of inducing early disclosure of inventions so as to avoid delays in commercialization and, in a model in which scope is defined as the number of applications of the patented technology covered by the patent claim, conclude that the relevant policy variable affecting disclosure is scope rather than length.

The second aspect has been addressed both in the context of models characterizing innovation as a two-stage process of “basic” and “applied” research and of “quality ladders” models in which innovation is represented as a sequence of improvements over an original invention (Scotchmer, 1999). In the first case, patent scope is understood as the primary factor affecting the relative bargaining power of the pioneer and the improver, with broad patents forcing second-generation innovators to obtain licenses from pioneers. The problem is to divide profits in a way that reflects the costs and the social value each invention generates. As Scotchmer (1991) observes, however, in order to provide sufficient incentives for applications, patent scope should be chosen so as to allow the developer to appropriate the entire social surplus of his innovation but doing so would jeopardize the incentives of the pioneer, because he would not be compensated for the positive externality his invention provides to follow-on innovators.

Green and Scotchmer (1995) make a case for broad patents to provide sufficient incentives to basic research, but also note that an increase in scope might be detrimental to the first innovator, as it increases the second innovator’s ability to threaten not to invest unless the first innovator shares investment costs\(^\text{11}\). Crucial to their argument is the possibility that first- and second-generation innovators can sign agreements ex ante, i.e. before the investment by the first

\(^{11}\) In addition to this, they show that optimal patent length should be longer when research and development are undertaken by different firms in order to restore innovators’ incentives that are reduced by a (necessarily) suboptimal division of profits.
innovator is sunk. Chang (1995) also argues in favour of broad patents for pioneer inventions stressing that broad patent protection should be extended in particular to those inventions that have little value on their own relative to improvements. However, his conclusion about the desirability of collusive agreement between pioneers and improvers are radically different from Green and Scotchmer’s in light of the incentives for “inventing around” that a lax policy towards cooperative agreements would create.

One important implication of extending patent scope can be appreciated by allowing for heterogeneity of the original innovations. For instance, Hopenhayn and Mitchell (1999) distinguish innovations by their potential to spur further lines of research, that is by their fertility. They suggest the implementation of a menu of combinations of patent length and scope that would induce revelation of privately held knowledge concerning innovations’ fertility.

A “quality ladder” model of innovation is developed by O’Donoghue, Scotchmer and Thisse (1998), who introduce the distinction between “lagging breadth” and “leading breadth”, where the latter is relevant to determine the “effective life” of a patent, i.e. the period of time between the grant of a patent and the moment in which the underlying innovation is displaced by a non-infringing improvement. They consider the choice between patents with infinite length and limited leading breadth and patents with limited length but infinite leading breadth to achieve a given rate of innovation and highlight the fact that the first policy is more effective in terms of the minimization of R&D costs, whereas the second policy limits market distortions because it implies a shorter effective life.

To sum up, the long way covered by economic analysis from the Nordhaus study of optimal patent design, constitutes a refinement of the perception that the optimal shape of the intellectual property system depends on a variety of factors. The optimal design of patents depends, among other things, on the nature of the
market structure, on the elasticity of demand and on the cost-reducing impact of innovations. As a consequence, it varies according to the industrial sector and the country considered. Unfortunately, the attempts to tailor the basic design of the patent system to the characteristics of the context considered may be rendered unfeasible by the costs such an attempt would pose. The perception that a “one-size-fits-all” system represents an imperfect compromise, however, should nurture the idea that other complementary policies should be devised in order to avoid some of the drawbacks it originates.

5.3. The role of IP enforcement and litigation

Eric von Hippel (1988) has argued that patent grants amount to no more than a right to sue others for infringement. The statement is striking in light of the fact that the majority of the theoretical analyses of the patent system assumes that patents provide perfect protection against imitation\(^{12}\), but has indeed some merit. Many aspects of the patent system contribute to confer to litigation a crucial role. The most relevant relates indeed to the very nature of intellectual objects, and particularly to the attribute of non-rivalry. Whereas in the case of tangibles the physical characteristics of the protected good clearly define the boundaries of the object of the right, in the case of intellectual property non-rivalry precludes a clear definition of the extent of legal protection at the time the right is granted. It follows that, differently from infringement of rights related to physical assets, whose definition can be clearly provided ex-ante, in the case of property rights over intangibles the definition of infringement is predominantly assigned to ex-post conflicts.

This implies that both the effective breadth of patents (interpreted as the range of product and processes covered by the patent) and the strength of patent protection (understood as the probability of winning an infringement case brought

\(^{12}\) Notable exceptions are Katz and Shapiro (1987) and Dasgupta (1988), who consider the issue of imitation in the context of patent races, and Gallini (1992), who explicitly models costly imitation.
to court) are actually determined by the legal system, rather than fixed ex-ante by patent law. In turn, this suggests that enforcement costs may significantly affect the extent and direction of IP-induced incentives to innovate\textsuperscript{13}.

Intellectual property litigation has increased dramatically in the past two decades. Lanjouw and Shankerman (2001) report that the number of patent suits rose by almost tenfold across the period 1978-1999, although the rise in litigation was matched by a comparable increase in patenting activity, so that suit filing rates have remained roughly constant. It is not clear to what extent this trend in litigation is related to recent patent system reform. Also not conclusive is the empirical evidence linking IP system reform to increased patenting activity (Kortum and Lerner, 1999; but see Hall and Ziedonis 2001 for a less cautious view). One piece of conclusive evidence concerns the distribution of average suit filing rates across technology sectors that is remarkably skewed, with computer, biotechnology and non-drug health patents having a substantially higher probability of being litigated than patents in other technology fields.

Both anecdotal and empirical evidence suggest that firms tend to enforce patents more aggressively (Hall and Ham, 1999), which is consistent with the already mentioned survey results indicating strategic objectives (both offensive and defensive) as the prominent motives for patenting (Cohen et al. 2000). In other words, IP litigation is increasingly perceived as a competitive weapon (Rivette and Kline, 2001).

In spite of the growing importance of IP litigation, the literature on the subject is still underdeveloped. While in the past decade increasing attention devoted to the issue of IP litigation has provided a rather satisfactory account of the phenomenon from an empirical perspective, theoretical analyses are still limited.

\textsuperscript{13} In addition to this, litigation often performs the relevant role of ensuring an ex-post form of control over patent validity, since the most common response to an infringement suit is a challenge to the validity of the contested patent. This is particularly important in light of the perceived decline in the quality of issued patents.
In the following, a brief account of both empirical and theoretical contributions on the subject will be offered, with a focus on strategic aspects of IP litigation.

### 5.3.1. Empirical evidence on patent litigation

Empirical analyses of patent litigation could be roughly divided into three research avenues: 

a) studies of the determinants of patent suits and the choice between litigation and settlement; 

b) studies of the impact of enforcement costs on the intensity of innovation, and 

c) studies of the impact of litigation on the nature and direction of innovation incentives.

Analyses sub (a) tend to confirm the theoretical findings of the law and economics literature (see Cooter and Rubinfeld, 1989 for a survey) and add some IP-specific results. As for litigation in other domains, three factors play a crucial role in the decision to file a suit.\(^\text{14}\) The first is the uncertainty or the asymmetric distribution of information over the outcome of the trial. Consistent with this hypothesis is the finding by Lerner (1995) and Lanjouw and Shankerman (2001) that litigation rates are substantially higher in new technology areas, and especially in biotechnology, where uncertainty is amplified by the limited past experience in litigation.

The second factor is given by the magnitude and distribution of the benefits from litigation. The probability of filing a suit is thus positively correlated both to the degree of stakes asymmetry (Siegelman and Waldfogel, 1999) and to patent breadth (Lerner, 1994). Both of these findings are coherent with the idea that successful suits bring benefits in terms of reputation. Lanjouw and Shankerman (2001) address more directly the role of reputation building in the decision to litigate by showing that the probability that corporate patentees resort to litigation is substantially higher when citations to the contested patent come from firms

\(^{14}\) The law and economics literature identifies an additional factor positively influencing the probability of litigation, namely the likelihood of infringement detection, but this factor is not central to empirical analyses of patent litigation.
active in closely related technological fields. In addition to this, they also provide evidence of the existence of what they call the “publicity effect”, namely an increase in citations of the contested patent shortly after the suit, which they take as evidence of the fact that competitors perceive the patentee’s willingness to aggressively enforce his rights.

Finally, the choice between litigation and settlement is affected by the relative costs of trial and settlement. The cost of trial is generally proxied by the identity of the patent-holder, namely by whether the patent-holder is individual or corporate, domestic or foreign. Lanjouw and Shankerman (2003) find that domestic patents are five times more likely to be litigated in all the technology areas they consider, which is consistent with the hypothesis that domestic individuals or firms bear lower litigation costs. They also find that patents owned by domestic and non-Japanese individuals are more likely to be litigated than patents owned by corporations in the same countries and interpret this finding as evidence of the fact that corporations face not only lower litigation costs but also lower settlement costs and that the latter effect tends to dominate the former.

Among the additional factors that have been found to have a bearing on the likelihood of recourse to trial, three are worth mentioning: firm dimension, the frequency of interactions among patentees and possession of a patent portfolio. Lanjouw and Shankerman (2001) interpret the positive correlation between the three mentioned variables and settlement rates as evidence of the importance of intellectual property “trading” (licensing and cross-licensing) and repeated interaction as mechanisms to promote patent dispute resolution.

As for the second branch of empirical literature, that explores the nexus between enforcement costs and intensity of innovation incentives, the literature identifies a positive relationship between patent value and the probability of litigation that in turn suggests that the dilution of incentives associated to costly enforcement of patents has the strongest incidence exactly on those innovations the patent system intends to promote (Lanjouw and Shankerman, 2003). Costly enforcement of IPRs affect the level of incentives they provide both directly and
indirectly. The direct effect depends on the burden of legal costs that must be netted out from any benefit patents potentially convey. The indirect effect follows from the reduction in patent value associated to the fact that, if enforcement is costly, a certain degree of infringement will be tolerated. Patent value is inferred from the number of claims (a proxy for breadth), the number of forward citations and the number of self-citations (a proxy for the extent to which a given patent is related to other technological activities pursued by the firm).

Lanjouw (1998) addresses various aspects of the impact of enforcement costs on innovation in an analytical setting that combines an infringement/litigation game into a patent renewal model. Among other things, her simulation results show that the reduction of patent incentives increases if litigation costs increase, if the statutory length of protection is reduced and if there is a shift from the “British” to the “American rule” of costs allocation.

Finally, the hypothesis that litigation costs tend to distort innovation incentives is confirmed by two types of results (Lerner, 1995; Lanjouw and Shankerman, 2001 and 2003). Firms facing higher litigation costs (and especially small firms) tend to a) direct research activity and patenting towards technology areas in which patenting by rival firms is limited or absent, and b) to avoid patenting in areas where firms with lower litigation costs (as proxied by the firm’s experience in patent litigation and by its paid-in capital) have previously patented.

5.3.2. Theoretical analyses of patent litigation

Litigation in the patent field differs from litigation in other areas in many respects (see Siegelman and Waldfogel 1999 for an empirical comparison of litigation in various areas). One of the most relevant distinguishing features is that the choice between litigation and settlement, as well as the final outcome of the trial, do not only affect the distribution of the value of the subject of litigation (that is, industry profits) but the value itself (Meurer, 1989). Indeed, settlement of
patent litigation often involves licensing, and thus a change in industry structure due to the entry of the alleged infringer. The issue of licensing is thus crucial to most of the analyses of IP litigation, although often with different implications.

For instance, Aoki and Hu (1999) consider a model of imperfect patent protection in which a patentee faces the choice between licensing and litigation in order to deter infringement. In their model, the legal system is captured by two variables, namely the legal costs of litigation and the probability of plaintiff victory. These two variables affect the patentee’s decision to licence in that they influence the credibility of the patentee’s threat of bringing suit to the infringer and the willingness of the latter to go to court. One of the most relevant results of Aoki and Hu’s model is that the socially optimal legal regime (one that achieves a balance between ex-post efficiency and ex-ante innovation incentives) is characterized by a monotonic relationship between the marginal cost of innovation and the strength of patent protection. This is because stronger protection makes credible the threat of litigation, thus allowing the patentee to deter infringement without recourse to licensing and therefore to appropriate higher profits than would be the case with licensing (monopoly profits rather than duopoly profits plus the licensing fee).

The relationship between strength of patent protection and patentee’s profit identified by Aoki and Hu (1999) is reversed by Llobet (2003) in a model that examines the effects of litigation in a setting characterized by cumulative innovation and private information about innovations’ value. In this setting, an important share of the patentholder’s profit is assumed to depend on the licensing fees obtained by follow-up innovators, along the lines of O’Donogue et al.(1998) “quality ladder” model. Strengthening patent protection has the double effect of increasing the level of licensing fees a patentee can extract from future innovators and of reducing the number of future innovators that are willing to enter and may thus ultimately be detrimental to the patentholder.

IP litigation strategies play a crucial role in affecting both exit and entry decisions. However, to our knowledge few formal models deal with strategic
aspects of intellectual property litigation. Meurer (1989), Waterson (1990) and Crampes and Langinier (2002) investigate the link between patent litigation and entry of potential competitors in a setting characterized by a single incumbent and a single potential entrant, whereas Choi (2009) considers a setting characterized by patent pools and cross-licensing.

Meurer (1989) analyzes patent licensing that is induced as part of a settlement agreement to avoid litigation on the issue of patent validity. Meurer’s take-it-or-leave-it bargaining model assumes that the patentholder possesses private information concerning patent validity and derives the existence of a “bluffing equilibrium” in which the patentholder bluffs about validity by refusing to license. In addition to this, Meurer explicitly considers the role of antitrust policy deriving somewhat counterintuitive results. A permissive antitrust policy is shown to have no effect on the probability of licensing, as valid patents are never licensed.

Waterson (1990) examines the influence that the characteristics of the patent system and of the enforcement framework exert on the selection of product variety by a potential entrant, assuming uncertainty about the extent of patent protection. A relative increase in the degree of appropriability patents afford is desirable, from a social welfare viewpoint, if consumers care little about variety, as incumbents’ incentive to invent is enhanced and the “business stealing effect” due to new entry is reduced.

Crampes and Langinier (2002) elaborate a formal model exploring the intensity of the monitoring effort a patentholder should make in order to supervise the market and react in case of infringement. They consider the strategic choice between litigation, settlement and accommodated entry and investigate the effect of the monitoring effort on the entry decision of potential competitors. Differently from Meurer (1989) and Waterson (1990), they assume that there is no uncertainty about the exact coverage of patent protection, although they also assume that patents provide only imperfect protection because of the need to detect infringement.
The assumption of uncertain patent protection characterizes also Choi’s 1998 model. The hypothesis crucial to the model is that patent litigation may reveal important information concerning patent validity. This informational externality constitutes a deterrent for the patentee from filing an infringement suit. Choi thus investigates how this externality affects the dynamics of entry and derives from his analysis implications for the determination of the optimal strength of patent protection. He shows, rather counter intuitively, that a strengthening of patent protection may be detrimental to the patentee, as the entry process may be a waiting or pre-emption game depending on IP strength and expected payoffs of the incumbent varies discontinuously according to the nature of the entry game.

Lanjouw and Lerner (2001) address directly the issue of predatory behavior in the context of patent litigation by studying both theoretically and empirically the use of preliminary injunctive relief in patent litigation. They evaluate the implications of this legal remedy for ex post efficiency and ex ante incentives, with a concern for the possibility that this legal instrument may be opportunistically used by large firms in order to impose unfavourable settlement terms to financially constrained smaller firms. Their empirical results show a positive relationship between plaintiffs size and the likelihood that they will request an injunction.

6. Concluding remarks

This paper has provided an overview of the main legal and economic perspectives on the public institutions for the protection of intellectual property rights. The bulk of the literature on innovation and intellectual property has focused on the design of the publicly devised aspects of the intellectual property system (including length and breadth, but also enforcement aspects), exploring ways to
address the trade-off between incentives and efficient access to intangible assets. The review of this literature has highlighted that the functions performed by the IP system as well as the optimal shape of IPRs depend on the technological features of the specific sectors considered. Clear-cut conclusions are difficult to reach on the basis of the present knowledge. Further complexity would be added by considering also the role played by private, informal institutions. This sort of institutions interacts with the publicly devised aspects analyzed in this paper, reshaping the intellectual property framework to ensure the creation of an institutional environment that better serves their purposes and, under some circumstances, a more efficient institutional environment, have been the object of considerably less attention. This has been the starting point for the research performed by the Siena Unit within the RefGov project.
References


Anticommmons in Biomedical Research", *Science* 280, pp.698-701.


