Economic theories about the benefits and costs of patents

Roberto Mazzoleni, Richard R. Nelson, *Journal of Economic Issues,* vol.32, n°4, 1998.

Since the early 1980s, patent policy in the United States has been strengthened, broadened, and extended to areas and actors where earlier patenting was relatively rare. The Court of Appeal for the Federal Circuit, established in 1982 to deal with patent litigation cases, has upheld a significantly higher proportion of district court decisions of patent validity than earlier was the case [Dunner 1988]. Increasingly, patents have been granted for inventions or discoveries that are a far distance from practical applications; this trend has been particularly evident in biotechnology. The Bayh-Dole Act of 1980 and successive legislative initiatives have effectively pushed universities and government laboratories to apply for patents on the results of government-sponsored research, whereas the earlier norm was to place such results in the public domain. The trend toward strengthening intellectual property rights has expanded to the international arena first through the GATT negotiations and now through the proceedings of the World Trade Organization [UNCTAD 1994]. These policy trends make it urgent to reflect upon some basic questions regarding the patent system.

What are the social benefits and costs of awarding patents for inventions? Many economists and patent lawyers seem to think that the answer to this question is simple and settled, at least theoretically. In this paper, we discover that the answer certainly is not simple and currently not well settled. There are a number of different theories that give different answers and only limited knowledge of where these different theories apply.

1. The anticipation of patents provides motivation for useful invention: we will call this the "invention motivation" theory.

2. Patents induce inventors to "disclose" their inventions when otherwise they would rely on secrecy, and in this and other ways facilitate wide knowledge about and use of inventions: we will call this the "invention dissemination" theory.

3. Patents on inventions induce the needed investments to develop and commercialize them: this we call the "induce commercialization" theory.

4. Patents enable the orderly exploration of broad prospects: we call this the "exploration control" theory.

Of course, these purposes are not necessarily mutually exclusive. To wit, the anticipation of a patent may stimulate the original invention, and the holding of a patent may prompt its subsequent commercialization. But certain versions of the theories are at odds; for example, one version of Theory 2 assumes that inventions will occur without patents and that the presence of patents mainly serves to widen use, which is quite the opposite of the most familiar version of Theory 1. F. M. Scherer reminds us (private communication) that, while Theory 1 is today the most familiar, Theories 2 and 3, like Theory 1, have been around for a long time. The argument that patents provide an inducement to disclosure is an old one. And the British Parliament extended James Watt's steam engine patent, a patent on an invention already made, under the argument that to do so would spur commercialization. Theory 4 is, however, relatively new and, as we will argue, opens up a Pandora's box of issues.

In general, the different theories about the benefits and costs of patents make different assumptions - although sometimes these are implicit, not explicit - about at least several of the following context conditions:

1. The nature and effectiveness of means other than patents to induce invention and related activities. These "other means" may be as diverse as government grants and contracts or strong first mover advantages.

2. Whether the group of potential inventors is likely to work on diverse and non-competing ideas, or whether the group is likely to be focused on a single alternative or a set of closely connected ones. Basically the issue here is whether or not more inventing input yields more useful inventing output or mainly duplication of effort and waste.

3. The deterrent effect of the presence of patents on unauthorized use of a technology and on the transaction costs involved in licensing an invention.

4. Whether the multiple steps in the invention, development, and commercialization of a new technology tend to proceed efficiently within a single organization, or whether efficiency is enhanced if different organizations are involved at different stages of the process.

5. What we will call the topography of technological advance, by which we mean the manner in which inventions are linked to each other temporally, and as systems in use.

At least some of these conditions are partly endogenous to the nature of the patent system. They are themselves influenced by the strength and scope of the patent protection within a field of technology. This is clearly so for context condition 2, where the extent to which there are invention races may depend on the kind of patent protection that inventors see as available. Condition 3 surely is partly endogenous, and condition 4, the extent to which it is efficient to involve more than one organization in inventing, surely depends on transaction costs that, in turn, may depend on the strength of patent protection. However, other factors, in addition to the patent system, influence the context conditions identified above. We believe that these factors are to a considerable extent exogenous to the characteristics of the patent system itself.

Each of the broad theories - differentiated in terms of the social purposes for patents on which they focus - tends to involve particular common assumptions about certain of these context conditions. However, in some cases the broad theories have several variants that differ in their context assumptions. As we will see, in such cases there may be overlap of the broad theories with, for example, a particular variant of Theory 1 looking a lot like a variant of Theory 3. In any case, the implications of the theories are very sensitive to the assumed context conditions.

Theory 1 - Patents Motivate Invention

Theory 1 - that patents motivate useful invention - unquestionably is the most familiar one. Indeed, much discussion about the benefits of patents proceeds as if motivating useful invention were the only social purpose served by patents, and that patents always serve this purpose productively. In many cases, neither presumption may be valid.

All versions of Theory 1 presume either that if there were no patent protection, there would be no invention, or, more generally, that absent a patent system, incentives for invention would be too weak to reflect the public interest (context condition 1) and that the prospect of a patent enhances invention incentives. Also, generally they assume that the stronger patent protection (in various senses), the more inventing will occur. Under what might be called the canonical version of Theory 1 - associated with the models of K.J. Arrow [1962], W.D. Nordhans [1969], and F.M. Scherer [1972] - it is assumed (generally implicitly) that the group of inventors is diverse and working on different and generally non-competing things (context condition 2). Thus, under this version of Theory 1, stronger patent protection results in a greater number of useful inventions or in better inventions.

Most articulations of Theory 1 presume, implicitly or explicitly, that the invention is used by the inventor through embodiment in a production process used by or a product sold by the inventor's firm. However, in his statement on the role of patents, Arrow does address the problem an inventor has in selling his invention to someone else in the absence of legal property rights on it. Arrow [1962], and more recently R. Merges [1995] and A. Arora and A. Gambardella [1994], have taken the position that strong property rights on an invention reduce the transaction costs of licensing it (context condition 3). Under this position, strong patents serve the purposes of Theory 1 by providing incentives to invent for parties who are limited in the extent to which they can employ the invention themselves (context condition 4) by facilitating the sale of rights to an invention. As we will see, this variant of Theory 1 overlaps certain versions of Theories 2 and 3.

In most versions of Theory 1, it is assumed, generally implicitly, that the social benefits of a particular invention are strictly its final use value (an implicit assumption about context condition 5). The social benefits of patent protection stem from the additional invention induced by the prospect of a patent. And the social costs of a patent are the restrictions on use associated with the monopoly power conferred by a patent. This formulation of Theory 1 leads naturally to the analysis of optimal patent strength, in the sense of duration [Nordhaus 1969; Scherer 1972] or breadth [Klemperer 1990] and in terms of a tradeoff between the increased inventing induced by greater patent strength and the increased costs to society associated with the stronger monopoly position of the patent holder [see also Gilbert and Shapiro 1990].

The variants of Theory 1 we have been considering up to now presume that inventing firms do not duplicate each other in their innovative projects and thus (context condition 2) that more inventive effort, and more inventors means more useful inventing. The theory takes on a different look if, instead, competition in R&D is allowed. When firms are presumed to be focused on a common research strategy or a common inventive goal, this gives rise to "invention race models" [Loury 1979; Dasgupta and Stiglitz 1980a]. If the presumed common focus is on a broader but still limited "pool" of invention prospects, one has the "overfishing" models [Barzel 1968; Dasgupta and Stiglitz 1980b].

There is some ambiguity within these theories as to whether racing or overfishing would occur in the absence of effective patent protection on achieved inventions and to what extent patents strengthen the incentives to achieve an invention first. The simplest interpretation is that, absent the lure of a patent, no one would try to invent (context condition 1). In this case, races and overfishing induced by the prospect of a patent may be better than nothing. Another interpretation is that some inventing would occur even if no patents were granted in the field, and there would be some racing as various parties sought to achieve a head start on a new product. Under these conditions, the anticipation that a patent will be awarded to the inventor that gets there first may make the racing problem worse, the more so the broader the scope of patent protection.

More generally, such models point to a number of reasons why the increase in total inventive effort induced by the lure of a patent no longer is an unambiguous plus. If inventors perceive that other inventors are in the game (i.e., if there is free entry/competition in R&D), they will see that their returns are dependent not simply on whether they achieve an invention, but on whether they achieve it first. Relative to the social optimum, patent protection will result in competitive R&D market equilibria where firms invest their resources at a faster rate, and too many firms race toward the same invention goal [Dasgupta and Stiglitz 1980a].

It would seem that a consequence of these kinds of invention inefficiencies induced by strong patents would be to shift the tradeoff between the benefits and costs of stronger patents so as to increase the latter. Thus, other things being equal, society ought to opt for stronger patents in fields where stronger intellectual property protection yields a larger flow of valuable inventions than in fields where stronger patents lead largely to more hounds barking up the same tree. And, indeed, this is the case in the model of optimum patent duration developed by D.G. McFetridge and M. Rafiquzzaman [1986].

The models considered thus far assume generally (for an exception, see Dasgupta and Stiglitz [1980b]) that the prospect of a patent induces inventive activity, even though some of such induced activity may be socially wasteful. The theoretical picture changes if one asks what happens after a patent is granted. In the simple race model, it tends to be assumed that, once one party achieves the invention, a patent is granted, and that ends the race. Since, under those assumptions, continuation of the race only would have led to more wasted effort, the ending of the race is an economic plus. On the other hand, if a patent is not awarded, while there may be duplicative inventive efforts, there may be more ultimate competition in the market where the invention is used. Under those circumstances, if the lure of a patent is not needed to induce inventive efforts in the first place, it is not clear that the granting of a strong patent yields positive net social value.

And what if the inventive efforts are not all aimed at exactly the same thing? Patent claims in general extend well beyond the particular specification described in detail [Merges and Nelson 1990]. Thus, the awarding of a broad patent to one party inventing in a field can cause other inventors to stop or divert their efforts, even if their inventions would have been somewhat different (context condition 3). J. Lerner [1995] has found exactly this effect in the field of biotechnology. In particular, the holding of patents by large firms tends to deter small firms from trying to invent in the same areas.

The threat of a suit clearly can dampen incentives to invent in a field. The credibility of the threat depends on how broadly the claims of existing patents are written, and how the courts treat infringement suits. As

P. Klemperer [1990] has argued, there is a tradeoff between granting and protecting broad claims, which at first glance would seem to increase the incentive to invent and patent, and the scaring away of other inventors after a broad patent has been granted.

The issue of the consequences of greater patent length or scope becomes even more complicated if today's inventions not only have direct use value, but also set the basis for subsequent inventive efforts (context condition 5). Arrow especially called attention to the possibility that the principal use of some inventions may be as inputs to further invention. T.W. van Dijk [1994] considers what he calls the "height" of a patent, by which he means the extent to which the patent controls subsequent improvements and variegations in the initial invention. S. Scotchmer, J. Green [Scotchmer 1991; Green and Scotchmer 1995], and others have dealt in some detail with the issue that, if the original inventor is able to capture only a small portion of the benefits from follow-on inventions, profit incentives do not adequately reflect the social interest in a "prospect opening" invention. On the other hand, the long-term effect of granting a broad, strong patent on the initial invention will depend on how the presence of that patent affects subsequent inventing. These considerations lead us into Theories 3 and 4, which we will treat later.

As the above discussion testifies, Theory 1, that patents provide the key motivation for invention, comes in a large variety of shapes and flavors. This is largely because the lion's share of the conceptual work on patents by economists has been oriented by Theory 1; there has been much less work on the other theories.

Relatedly, virtually all of the empirical work exploring the effect of patents have been guided by Theory 1. There have been three major relatively recent interview or survey studies that have explored the perceived importance of patents as a means of enabling firms to profit from their inventions, all of which have explored inter-industry differences. These include a study by E. Mansfield [1986], what has come to be called the Yale survey [Levin et al. 1987], and what we will call the Carnegie Mellon study [Cohen, Nelson, and Walsh 1996]. All of these studies come to basically the same conclusion, which is partially about the efficacy of patents and partly about the effectiveness of other means to enable firms to profit from their innovations - context condition 1. In a nutshell, patents are an important inducement to invention in only a few industries.

And this conclusion holds for most industries where firms do a lot of R&D. In pharmaceuticals, patents do seem to be an important part of the inducement for R&D. However, in industries like semiconductors and computers, the advantages that came with a head start, including setting up production, sales, and service structures and moving down the learning curve, were judged much more effective than patents. In some of these industries, the respondents said that imitation was innately time consuming and costly, even if there were no patent protection. In others, it was said that technology was moving so fast that patents were pointless.

We will discuss later some of the limitations of these studies, even as they bear on Theory 1. However, at the present time they provide most of the systematic empirical information that we have regarding the benefits and costs of the patent system.

In any case, empirical studies of the sort sketched above have led a number of economists to downplay the relevance of patents, or at least express skepticism regarding the net social benefits of patent protection [see, e.g., Scherer and Ross 1990]. If the granting of patents is not necessary to induce inventions and adds little to overall incentives for invention, why incur the social costs of granting inventors a legal monopoly? If it is recognized that patents are necessary to induce inventing in a few areas but not in most, might it not be worthwhile to try to limit the domain where we grant patents? However, before leaning too far in this direction, it seems wise to consider other theories that cast the benefits and costs of patents in a somewhat different light.

Theory 2 - Patents Induce Disclosure and Wide Use of Inventions

Theory 2 has been part of the conventional wisdom of the patent policy community for a long time, although until recently it was poorly received by academic economists [Machlup 1958]. Under Theory 1, patents are necessary to induce inventing, but at the cost of restricting use. Under Theory 2, patents are not necessary to induce invention. Rather, patents encourage disclosure and, more generally, provide a vehicle for a quick and wide diffusion of the technical information underlying new inventions. That is, Theory 2 in effect turns Theory 1 on its head, or so it appears at first glance. That is not completely so since, in most specific versions of Theory 2, the possibility to make more profit through wider disclosure enhances the incentives for invention in the first place. However, in Theory 2, patenting does not cause the reduction in use of an invention that it does in the canonical version of Theory 1; indeed, the contrary is true, and the focus in Theory 2 is on incentives for disclosure.

The conventional version of Theory 2 assumes that an inventor can appropriate some returns from a new process or product simply by using or producing it while keeping the relevant information secret to prevent rapid imitation. The possibility of patenting the invention, however, lures the inventor into making public the relevant information. In earlier years, the argument was often couched in terms of society's access to the technology after the inventor had died [Machlup 1958]. However, in the modern world where companies, rather than individuals, are largely the custodians of invention-specific technological knowledge, the issue clearly must be posed more generally in terms of the speed, breadth, and completeness of information disclosure or leakage.

In our view, Theory 2 becomes interesting when it is assumed that the inventor, by him or herself, cannot exploit all possible uses of the invention. Then, to the extent that the holding of a patent facilitates licensing (context condition 3), this not only increases the rewards to inventing, but also facilitates wider use. The argument here is obviously kin to that in the version of Theory 1 that focused on how patenting might induce inventive effort from parties who cannot themselves directly use an invention. However, the focus here is on how patenting might extend use, rather than how it enhances, incentives for invention in the first place.

This argument would seem potentially important in contexts where secrecy can be effective in enabling an inventor to reap at least some returns. The Yale survey [Levin et al. 1987] suggests that this is the case for many process inventions. Various studies have shown that in certain industries firms customarily engage in general cross licensing of process technology, a sharing of technology that likely would be much more difficult if patents were not available on process technology.

While secrecy in general is less effective as a means of appropriating returns from product invention, we noted above that in many industries firms can profit from their product inventions without a patent. However, possession of a patent, rather than simple reliance on non-patent measures to reap returns, may make a firm more willing to advertise its inventions and to give technological information and assistance to a non-competing firm to help it adapt the invention to its own uses.

This interpretation obviously raises the question of how wide a range of use coverage a patent ought to grant. If one is focused on Theory 2, one is led to ask whether the presence of a broad, strong patent on an invention that has many potential uses, some of which the inventor may not be able to exploit directly, will facilitate wide use through licensing, or deter it by monopoly pricing backed up by threat of suit. Of course, if the analytic focus is Theory 1, the issue of incentives for the original invention becomes central [see, for example, Matutes et al. 1996].

In our view, the most interesting cases here involve inventions that need further work to tailor them to particular uses, and where one or more organizations need to be involved for that work to proceed effectively (context conditions 2 and 4). This gets us into Theories 3 and 4.

Theory 3 - Patents Induce the Development and Commercialization of Inventions

Theory 3, that the holding of a patent induces its commercialization, has received more attention recently than Theory 2 but still far less than Theory 1. While it has been part of the conventional wisdom for a long time (as noted, Watt's steam engine patent was extended to give him more time to effect commercialization), what brought Theory 3 to current attention undoubtedly was the argumentation that led to the passage of the Bayh-Dole Act in 1980. More on this later.

In its simplest version, Theory 3 would seem to be a variant of Theory 1, but with patenting occurring early in the process of inventing and with a lot of follow-on work needing to be done before the crude "invention" is ready for actual use. A patent at an early stage is seen as providing the assurance that, if development is technologically successful, its economic rewards can be appropriated, thus inducing a positive development decision. R. Eisenberg has called our attention to a variant or a supplement to this argument. It is that the possession of a patent enables the patent holder to go to capital markets to get development financing [Eisenberg 1997]. This capability might be important for a small firm facing large development costs before it can get its invention to market.

Theory 3 becomes distinctively different from at least the standard version of Theory 1 in circumstances (context condition 4) where one organization does the early inventing work but is not in a position to do the development work. Under this variant, the possession of a patent by the original inventor facilitates handing off the task to an organization better situated for development and commercialization. Years ago, W.F. Mueller [1962] pointed out that a large share of DuPont's product innovations were based on inventions bought from smaller firms. Similarly, in the 1920s General Electric bought and developed many inventions originally made by private inventors or small firms [Reich 1985].

As with Theory 2, many versions of Theory 3 are connected to a Theory 1 argument. Thus, if a first-stage inventor is in the game for profit and knows that profiting will require handing-off the invention to another organization for development, then expectation of a patent may be necessary to induce that initial inventing. But the emphasis in Theory 3 is on the facilitation of the hand-off.

As noted, Theory 3 was brought to the fore in the discussions that led to the Bayh-Dole Act, which, among other things, gave universities the patent rights on inventions that emanated from their government-funded research projects. Under Theory 1, which presumes that while patents may be needed to induce inventing, they should not be granted if inventing would go on in any case, the Bayh-Dole Act does not make any sense. The argument for the policy was that, while the inventions had been achieved using public monies, they would serve no economic purpose until they were developed to a point where they were commercial, and only companies had the capabilities to undertake such development (thus, there was separation of the locus of inventing from the locus of development - context condition 4).

Under the version of Theory 3 most clearly articulated in these discussions, a company would be unlikely to engage in development of a university invention unless it had proprietary rights. If the universities held strong patent rights, they would be in a position to sell such licenses (context condition 3). In contrast, if there were no patents, or if the government held them with a commitment to non-exclusive licensing, companies would be unlikely to invest in the necessary development work. This particular argument for the Bayh-Dole Act seems to presume that patents cannot be taken out on development work, or that the results of such development cannot be made proprietary in other ways. In many areas, patents do in fact emanate from development. Further, the existing studies [Levin et al. 1987; Mansfield 1986; Cohen et al. 1996] indicate that in many industries patents are not needed to induce development. A simple head start on commercialization can yield large profits on a new product, and secrecy often can protect effectively new process technology used by the developer.

Recognition of these facts has led some observers to be skeptical about the value of the Bayh-Dole reforms. However, if the licensee is a small firm that must marshal outside funds and may be swamped by quick imitation from a large firm, the case for the reforms might appear stronger.

Another interpretation of Theory 3 is that the possession of a patent gives the original patent holding organization - a university or small firm - incentive to push out its inventions to firms that can develop and commercialize them. This is basically an extension of one version of Theory 2 that we discussed above. It is a view of Theory 3 different from the one that implicitly denies that development work will lead to profitable product or process innovation for the firm doing it, absent a strong patent initially.

On the other hand, a case can certainly be made that, for many university "inventions" that were funded with public monies, the policy implications of Theory 1 - that if one does not need to grant a patent to get an invention, one should not grant a patent - are basically correct. The results of research would be published in any case. Firms, in many instances, would have ample incentive to work with and "develop" what comes out of university research. They usually can patent the developments or gain the advantage of a head start of the market, or both (context condition 1). No ex ante grant of an exclusive license is needed to motivate this work, and the presence of a patent and the requirement to get a license to do further work on the original idea may restrict the number of parties who will do that work.

This argument is particularly strong, it seems to us, if potential developers are diverse in the directions they might follow (context condition 2) and if licensing arrangements on preliminary ideas, whose ultimate commercial value is unclear, are not easy to work out (context condition 3). On the other hand, the holding of a patent may provide motivation for the inventor to seek out and work with a variety of different potential developers.

Theory 4 - Patents Enable Orderly Development of Broad Prospects

An implicit feature of most versions of Theory 3 is that, while significant resources and risk taking may be needed to develop an invention, there is basically one commercial product at the end of the rainbow. The prospect theory - Theory 4 - differs from Theory 3 in that an initial discovery or invention is seen as opening up a whole range of follow-on developments or inventions. We note that many university "inventions" are of this sort.

Under Theory 4, the holding of a broad patent on a prospect opening invention permits the development of the full range of possibilities to proceed in an orderly fashion. Under E.W. Kitch's articulation [1977], unless there is a broad patent on a prospect opening invention, development of the prospect is likely to proceed in a wasteful way (in contrast, under Theory 3 without patent protection on the seed invention, development will not proceed at all). Under the "exploration control" theory, the problem is that, unless there is a controlling patent, a lot of people see the same things (context condition 2) and know that their competitors also see them, and the consequence will be races for specific targets of opportunity and general overfishing in the prospect pond. Thus, a broad patent on the initial invention is necessary if the "wasteful mining of the prospect" or the "overfishing of the poop is to be avoided.

As one of us has previously argued [Merges and Nelson 1990], one can come to a very different view of the benefits and costs of giving a wide patent on a "prospect opening" invention if one assumes (context condition 2) that different inventors see very different things in the prospect and would do different things (as in the Arrow, Nordhaus, and Scherer theory). Indeed, under this version of the "prospect theory," there might be very high social costs to granting a broad initial patent that gives monopoly rights on the exploration of the prospect. This would cut down on the number of diverse inventors who would be induced to work on the prospect in anticipation of a profitable invention down the road, since their ability to work that invention would be constrained by their ability to negotiate a license with the holder of the original prospect defining patent.

This argument suggests that an important issue on which the analysis of the benefits and costs of granting patents on broad prospects turns is what one understands about the market for patent licenses (context condition 3). If one assumes that, in general, potential licensees and patent holders have little difficulty in reaching a license agreement (that is, that the transaction costs of patent licensing are small), then one may take a relatively relaxed view of the costs of granting a large prospect controlling patent, even when one believes that potential explorers of the prospect are diverse in terms of what they would do. On the other hand, if one believes that transaction costs often are high, and patent holders are prone to litigation, one is less sanguine about this.

We note here that most of the recent writing by economists on the role of patents in prospect opening inventions or, more generally, inventions that set the stage for a number of follow-on inventions, have been motivated by Theory 1 questions, focused on the effect of different kinds of patent protection on incentives for the initial invention [Scotchmer and Green 1990; Scotchmer 1991; Green and Scotchmer 1995; Chang 1995]. A basic issue addressed in these papers is that the incentive to make the initial invention, if expected profit is the driver, depends on the extent to which the creator of the initial invention is able to share in the returns to follow-on inventing. In general, it is assumed that the initial inventor is not in a good position to make the follow-ons; hence, the inventor must reap returns on these through license revenues or through other kinds of profit-sharing agreements. The writers generally stress that, if potential follow-on inventors have to take out a license after the fact of their inventing in order to exploit what they have achieved, then they are subject to hold-up bargaining by the initial patent holder, and if they expect this, they may be deterred from the effort in the first place. In some papers, ex ante licensing (that is, an agreement between the original inventor and the potential follow-on inventor written before the latter makes the investment) is seen as a possible solution.

In fact, most of the bargaining about licenses comes after, not before, a follow-on invention is made. Scholars like Kitch seem optimistic that bargaining will usually go smoothly. Our interpretation of the historical record leaves us less optimistic. The cases of the Selden and Wright patents in the fields of, respectively, automobile and aircraft technology indicate that - with respect to at least a few important technological developments - broad definition of pioneer patents has led to litigation and likely has forestalled the pace of technical advance [Merges and Nelson 1990]. More generally, the available evidence suggests that the transaction costs of technology transfer and licensing are usually considerable [Caves et al. 1983; Contractor 1981].

The issues raised by Theory 4 are particularly salient, we believe, in two kinds of contexts. One of these is when technological advances within a prospect are strongly connected (context condition 5) as they are in what Merges and Nelson [1990] call cumulative systems technologies. In such contexts, advancing a technology often will require the ability to use a number of already developed components, and hence it will necessitate either the ability easily to negotiate a license or an environment where litigation is not a serious threat.

The other setting in which Theory 4 issues are important is when an initial invention or discovery is a far distance from practical application, and its principal value is in providing clues as to how to proceed. In such cases, technological progress is likely to be furthered if a number of different parties follow the leads as they see them. But at the same time, it may be virtually impossible for any party to estimate with any confidence the expected value of taking out a license to follow those clues.

In any case, whenever an invention is understood as contributing to further invention potential, as well as creating a new or improved product or process of immediately final use, a question can be raised as to whether strong patents enhance or hinder technical advance over the long run. The question of how strong a patent should be, or whether a patent should be granted at all, no longer turns on analysis of a tradeoff between the positive effects on inventing of stronger patents and the restrictions in use of technology associated with a regime of strong patents, as in Theory 1. Rather, a good part of the argument is about whether the long-term effect on inventing of strong patents is positive or negative.

Let us sum up. The burden of this section is that there are a number of different theories about the benefits and costs of patents. We have tried to catalog them in Table 1.

Where Do These Theories Apply?

The proposition we now want strongly to espouse is that the appropriate question about these diverse theories is not "Which theory is the correct one?" but rather, "Where do the different theories apply?" Our attempt to specify the context conditions that seem to be presumed by the different theories and their variants is a first step toward identifying the domains of applicability. However, to go from specification of abstract context conditions to specification of particular industries and technologies requires a lot of detailed empirical work, most of which has yet to be done.

We undoubtedly know more about the domain of application of Theory 1 - that the expectation of a patent provides strong motivation for inventive activity - than we know about the applicability domain of the other theories. But even here, if the analysis of the previous section is on the mark, there is a lot we do not know.

Virtually all the systematic empirical work that has been done on the effects of patents has been guided by Theory 1, but it has been implemented on only one of the variants of Theory 1. All of the studies we briefly described earlier have gathered their data from firms with R&D laboratories. From those studies, we know (context condition 1) that in many industries, for firms of this type, it is possible to reap significant returns from industrial innovation without the aid of patent protection. However, to a large extent the effectiveness of the alternative mechanisms for reaping returns from inventing (for example, reaping the benefits of a head start through building a sales and service operation) involves the inventor in the production and sale (or use) of the product (or process) in question. Thus, these studies do not give us an effective window to explore the extent to which the prospect of patents induces firms or other organizations, with no way themselves to use an invention, to try to invent anyway in the hope of profiting through licensing.

Table 1. A Catalog of Patent Theories

Variants	*Issues*
Theory 1: Patents Motivate Inventing	
A. More inventing is better i. Inventors as a group do more inventing	In many industries, the prospect of a patent does not seem significantly to increase for inventing incentives
ii. New inventors are drawn in who cannot themselves directly exploit their inventions B. More inventing may be worse (patent races)	
Theory 2: Patents Encourage Wider Use of Inventions	
A. Patents induce disclosure rather than secrecy	But patents also enable the patent holder to restrict use
B. Patents induce licensing of inventions for uses the inventor cannot directly exploit - a close relative of Theory 1-A (ii)	
Theory 3: Patents Induce Development and Commercialization of Inventions	
A. A variant or an extension of Theory 1 with the patent coming early in the process i. Induces and enables development to be funded	But many inventions are developed and commercialized without a patent
ii. Stops invention races	University patents may restrict access to science and technology that otherwise would be in the public domain
B. Enables and induces transfer of an invention to another organization better suited to develop it i. A variant of Theory 1-A (ii) and 2-B ii. University transfer to firms: * Firms would not commercialize unless they control a patent, as in 3A * Induces academics to be entrepreneurs	
Theory 4: Patents Enable Orderly Prospect Development	
A. Absent a controlling patent, there will be races	But the presence and enforcement of a broad patent limit the parties who have motivation to work a prospect

The recent Carnegie Mellon survey does enable the analyst to make a distinction between large and small firms within an industry. A preliminary analysis suggests that in many industries small firms employ the licensing mechanism more than do large firms. But this is inconclusive and tells us nothing about whether the possibility of licensing induces firms and other parties outside of an industry to work to create technology useful in that industry [Cohen et al. 1996].

To our knowledge, there has been no systematic empirical work attempting to map out the relevance domain of Theory 2 - that patents induce disclosure of inventions and otherwise (for example, through licensing) facilitate the spread of an invention to uses that the inventing firm would not have made on its own. While there are many anecdotal examples of the latter, there has been no empirical study broadly guided by Theory 2.

Similarly, the information we have about the relevance domain of Theory 3, that the holding of a patent facilitates the development and commercialization of embryonic new inventions, is almost exclusively anecdotal. There are myriad examples of products that were initially invented by one party, often a private inventor or a small firm, that subsequently were developed and commercialized by another, often a large firm. However, we know nothing systematically about the industries and kinds of inventions where this is important or of the role of patents in such technology transfer.

Interestingly, we are beginning to get systematic information about just where the version of Theory 3 that focuses on the development and commercialization of university inventions is relevant [Gelijns et al. 1996]. The recent surge of university inventing has largely been in the biomedical area (particularly biotechnology), electronics, and software [Henderson et al. 1995]. This is also where most of the lucrative university licenses are concentrated.

Earlier we signaled that, in our view, Theory 4 - that the existence and use of a patent facilitates the orderly exploration of broad prospects - is the most controversial of the various theories about the social value of patents; we believe that, as a general rule, it is a hindrance, not a help, to technological advance to police the number and identity of parties who are working to advance a particular area of technology. As we noted above, one's evaluation of Theory 4 is intimately wrapped up with one's views on whether different inventors in a field are likely to be trying virtually the same thing, or whether they likely will be following different paths. Thus, the issue here is connected to the divide regarding Theory 1 about whether or not "invention races" are common. We believe that they are very uncommon, and we would cite as general evidence the very small number of applications to the U.S. Patent Office that result in interference proceedings: less than half of 1 percent. And a very large share of these are concentrated in a small number of technology classes (U.S. Patent Interference Records; numbers communicated by William Kingston).

This brings us to the following propositions regarding fruitful directions for empirical research to assess the relevance domains of the different theories about the social value of patents. It surely would be useful to develop some questionnaires that, following in the established tradition, explore the importance in different industries of the different theories, with particular emphasis on those aspects of Theory 1 that have not been explored, and on Theories 2-4. However, another way of cutting into the question is to recognize that there are certain broad classes of technology that differ in terms of the topology and dynamics of technological advance.

Much of the standard thinking about the role of patents presumes, explicitly or implicitly, that inventions come largely from ideas in the heads of their creators and stand separate from each other. For some inventions, these may be reasonable assumptions. Gillette's safety razor is not a bad example. But in many industries, either or both of these presumptions is far off the mark.

Patents would appear to play special kinds of roles and have particular kinds of impacts in science-based technologies, where by that term we mean ones in which the ability to create new products and processes is strongly influenced by a continuing flow of new scientific understandings and techniques largely emanating from university research. We note that not many technologies, or industries, are science based in the above sense. Both the Yale Survey and the Carnegie Mellon Survey contained questions about the importance of university research to technological advance in a line of business, as well as questions about the importance of patents. Both surveys yielded similar findings. Only a few industries rated university research as very important to them. Those that did clustered in three areas: medicine and other health-related products, electronics, and agriculture related (where clearly the reference was to university research funded by the Department of Agriculture). These industries clearly differ in important respects. However, particularly the first two clusters share certain things in common, we believe, that are relevant to this essay.

First, while we believe that patent races are quite rare, they are much more common in science-based industries than elsewhere. In these industries, multiple inventors not only see the same broad unmet needs, but at the same time also often pick up knowledge of research advances that suggest particular avenues to follow. Modern biotechnology is perhaps an extreme case in point. But in other areas, like superconductivity, particular scientific advances also often attract many players to the same opportunity.

In this kind of context, there is clearly a tension. On the one hand, under some conditions, and particularly where there is no good way of protecting the follow-on work, patent protection may be necessary to motivate investment of private resources in the development of commercial products based on scientific knowledge in the public domain. Absent such protection, it is quite possible that private investment will be thwarted became of the weak appropriability conditions. This is of course a Theory 3 argument, and we argued above that we doubt it is widely applicable. We have suggested that, in a wide range of circumstances, the advantages conferred by a head start, or the possibility of getting a patent on the product or process that is based on the original advance, or both, seem to provide ample incentive for the follow-on work. In such cases, the award of a patent on the initial discovery may unnecessarily restrict competition in the field.

Second, science-based industries are the arena where the need for "technology transfer" between university and industry researchers is most salient. We argued earlier that the popular case for giving universities strong patents - that companies will not develop these inventions unless they have an exclusive license - does not appear to be broadly valid, although it does seem salient in cases where the innovating firm is a start up. Also, it does appear to be true that, when universities are able to obtain strong patents and university inventors share in the financial rewards, incentive is provided for university researchers to reach out to industry. In a number of instances, university researchers, in control of patents, have been directly instrumental in the setting up of new companies. In these cases, patents (and the exclusive licenses granted on them by the university) may be essential to raising venture capital and to reducing transaction costs in licensing.

Third, science-based technologies are prominent among those in which an idea or finding may still need a lot of work to be brought to practice, and there may be significant uncertainty as to how that should be done. An early patent on such findings can narrow significantly the range of parties who have the incentive to do the follow-on work. The patenting of gene fragments is a striking case in point, but there are many others. Thus, particularly if patents are given on promising ideas that still are a long way from practice, the issues posed by Theory 4 are prominent in science-based technologies. Just when are broad patents useful to achieve a better coordinated mining of the prospect? When are they likely to hinder creativity?

Another broad class of technologies where patents appear to play distinctive roles is that of cumulative system technologies. By that term, we mean ones in which the operative products define a system involving a number of different components integrated into a system architecture. The system changes over time as the components and the architecture change. Examples include aircraft, automobiles, computers, and telecommunication systems.

The advance of a cumulative system technology often involves a number of different kinds of firms, in different industries, producing different components, along with (usually) firms that assemble systems. While the latter tend to be large firms, the former may be quite small. On a few occasions in the past, one organization established broad control, through intellectual property rights or other mechanisms, over the advance of a system technology. AT&T (from the 1920s until 1980) and IBM (from the 1960s until 1980) are good examples. However, for the most part, many different parties are involved in the advance of system technologies.

In these technologies, the development of standards, rather than the effective use of a portfolio of patents, has been the principal vehicle for achieving coordination. And while, in many cases, the prospect of a patent seems to have been important in inducing inventive activity, in other cases the holding of strong patents on key components by certain parties, or on a broad systems design concept, has resulted in litigation that has held up further development of the technology. The famous case of the Wright brothers aircraft configuration patent, and the tangle of patents that was characteristic of the early radio industry, are cases in point [Merges and Nelson 1990]. In both instances, the impasse was broken only by the establishment of a system of virtually universal cross licensing. In the post-World War II era, the development of both semiconductor technology and computer technology proceeded under a regime of relatively easy cross licensing and (after a few major early cases were settled) relatively little litigation.

Issues in Patent Reform

In this concluding section, we discuss selectively a few major current controversies regarding the benefits and costs of granting patents. Our principal purpose is to highlight that a good deal of each controversy is about what theory, or what version of a theory, is appropriate to the context, but that this issue often is glossed over. Just as almost all empirical work on the role of patents has been oriented by Theory 1, almost all patent policy issues are argued out on these terms, at least initially. However, it would seem apparent that many of today's salient patent policy issues are not adequately viewed through a simple Theory 1 lens.

Thus, consider the debate in the late 1970s that led to the Bayh-Dole Act. As we have noted, the argument that carried the day was that patents were required if inventions already achieved under federal funding were to get developed and commercialized. This is a Theory 3 argument, with a touch of Theory 2 as well. We have argued that the particular version of Theory 3 put forth most vigorously in these debates - that companies would not develop an invention unless they had a prior patent on it - probably was not widely relevant. On the other hand, Bayh-Dole undoubtedly has led to a significant increase in university entrepreneuring. Whether this is good or bad is a complicated question. However, the evaluation of Bayh-Dole, like the arguments that lay behind its genesis, must be in terms of Theory 2, 3, and 4 arguments, more than Theory 1 ones.

The same is true regarding the issue about patenting the codes on gene fragments identified in the human genome project. Here too, at least in the early days of the discussion, the issue was not that the prospect of a patent was needed to get the research work done; the research work was being funded by the government. Rather, the argument was that patents on the coded gene fragments were needed if companies were to be induced to use that information to achieve commercial products. That is, the arguments were those of Theory 3 and perhaps of Theory 4. We, like R. Eisenberg [1995], have questioned the validity of this theory in this particular context.

As matters have turned out, the belief that codes for gene fragments will be patentable has led to the birth of private for-profit firms whose business is to discover these codes in anticipation of profiting from licensing to larger companies who would take on the subsequent development work. Interestingly, several large pharmaceutical companies have argued that gene fragments should not be patented, but rather that their identified codings should be in the public domain. Their case is that the work in going from coded gene fragments to useful final products will be more costly if gene fragment codes are patented than if they were in the public domain. That is, they are arguing that the standard version of Theory 3 has it backwards. And several of these companies are putting their money where their mouths are, and supporting research to identify gene codes, with the condition that information be put in the public domain. The public policy issues here clearly are very complex. Only a few of them can be seen adequately through a Theory 1 lens, however.

The argument about intellectual property rights on software also is a complicated one. Some of the large companies advocating patent or patent-like protection on software pose their cases in terms of a simple Theory 1 argument. However, smaller companies and university researchers argue that there is little need for stronger intellectual property rights protection in order to induce software development and that patenting key elements of "software systems" will significantly cut down on the number of parties who will find it worthwhile to develop new software that is connected into those systems. That is, the issue is a Theory 4 one at least as much as a Theory 1 one. The case of those who argue against patents on software is that strong, wide intellectual property rights will hinder technical advance in this field. This issue recently has been broadly discussed in a symposium on intellectual property rights in software [Samuelson et al. 1994].

The three issues sketched above differ in important ways. However, they also have important commonalities and raise common questions.

Perhaps the most basic question they raise is whether the prospect ex ante of a patent, together with its ex post presence, stimulates or interferes with technical advance in a field. Under Theory 1, this question is an oxymoron. But under a more complex theory, the answer is not always apparent. The question, then, raises a serious issue about the appropriate domain of patents. We believe that this issue has been under the rug too long, and today it is badly in need of open examination. The argument that strong intellectual property rights in a field may smother technical progress is, of course, connected to assumptions about several of the context conditions discussed earlier. The following three questions seem of particular importance.

First, in what areas of technology are technical advances so strongly connected one to another, either temporally or in a system of use, that effective inventing today requires access to prior inventions? Second, what are the fields of inventing where progress generally requires the effective interaction of a number of different organizations? And third, do patents in fact contribute to, or hinder, the access and cooperation needed for technical advance in such contexts?

As indicated earlier, there currently is very little empirical research aimed at this cluster of questions. Our lack of knowledge here clearly limits our ability to analyze intelligently the current pressing issues of patent reform.

References

Arora, A. "Appropriating Rents from Innovation: Patents, Licensing, and Market Structure in the Chemical Industry." Unpublished manuscript, School of Industrial Administration, Carnegie Mellon University, 1995.

Arora, A., and A. Gambardella. "The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labor." Research Policy 23 (September 1994): 523-532.

Arrow, K. J. "Economic Welfare and the Allocation of Resources for Invention." In The Rate and Direction of Inventive Activity, edited by R. R. Nelson. New York: Princeton University Press, 1962.

Barzel, Y. "Optimal Timing of Innovation." Review of Economics and Statistics 50 (August 1968): 348-355.

Caves, R., H. Crookell, and J. P. Killing. "The Imperfect Market for Technology Licenses." Oxford Bulletin of Economics & Statistics 45 (1983): 260-262.

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

Cohen, W., R. R. Nelson, and J. Walsh. "University Research and Industrial Innovation: Findings from a New Survey." Unpublished manuscript, 1996.

Contractor, F. International Technology Licensing: Compensation, Costs and Negotiation. Lexington, Mass.: Lexington Books, 1981.

Dasgupta, P., and J. E. Stiglitz. "Uncertainty, Industrial Structure and the Speed of R&D." Bell Journal of Economics 11 (1980a): 1-28.

-----. "Industrial Structure and the Nature of Innovative Activity." Economic Journal 90 (June 1980b): 266-293.

Dunner, D. R. Special Committee on CAFC, American Bar Association, Section of Patent, Trademark, and Copyright Law, Annual Report, 1988: 314-325.

Eisenberg, R. "Public Research and Private Development: Patents and Technology Transfer in the Human Genome Project." Unpublished manuscript, University of Michigan Law School, 1995.

-----. "Patenting Research Tools and the Law." In Intellectual Property Rights and Research Tools in Molecular Biology, edited by the National Research Council. Washington, D.C.: National Academy Press, 1997.

Gelijns, A., R. R. Nelson, H. Raider, and B. Sampat. "A Preliminary Report on Columbia Inventing." Unpublished manuscript, 1996.

Gilbert, R., and C. Shapiro. "Optimal Patent Length and Breadth." RAND Journal of Economics 21 (Spring 1990): 106-112.

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

Henderson, R., A. B. Jaffe, and M. Trajtenberg. "The Bayh-Dole Act and Trends in University Patenting 1965-1988." Center for Economic Policy Research, Discussion Paper no. 433, 1995.

Kitch, E. W. "The Nature and Function of the Patent System." Journal of Law and Economics 20 (October 1977): 265-290.

Klemperer, P. "How Broad Should the Scope of Patent Protection Be?" RAND Journal of Economics 21 (Spring 1990): 113-130.

Lerner, J. "Parenting in the Shadow of Competition." Journal of Law and Economics 38 (October 1995): 463-495.

Levin, R. C., A. K. Klevorick, R. R. Nelson, and S. G. Winter. "Appropriating the Returns from Industrial Research and Development." Brookings Papers on Economic Activity 3 (1987): 783-820.

Loury, G. L. "Market Structure and Innovation." Quarterly Journal of Economics 93 (August 1979): 395-410.

Machlup, F. An Economic Review of the Patent System. Study of the Subcommittee on Patents, Trademarks, and Copyrights of the Committee on the Judiciary, U.S. Senate, 85th Congress, Washington, D.C.: Government Printing Office, 1958.

Mansfield, E. "Patents and Innovation: An Empirical Study." Management Science 32 (February 1986): 173-181.

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

McFetridge, D. G., and M. Rafiquzzaman. "The Scope and Duration of the Patent Right and the Nature of Research Rivalry." In The Economics of Patents and Copyrights, edited by J. Palmer, 91-120. Research in Law and Economics, vol. 8. Greenwich, Conn.: JAI Press, 1986.

Merges, R. "Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents." Tennessee Law Review 62 (Fall 1994): 75-106.

-----. "Contracting into Liability Rules: Institutions Supporting Transactions and Intellectual Property Rights." Unpublished manuscript, University of California at Berkeley, 1995.

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

Mueller, W. F. "The Origins of the Basic Inventions Underlying DuPont's Major Product and Process Innovations." In The Rate and Direction of Inventive Activity, edited by R. R. Nelson. New York: Princeton University Press, 1962.

Nordhaus, W. D. Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change. Cambridge, Mass.: MIT Press, 1969.

Reich, L. S. The Making of American Industrial Research: Science and Business at GE and Bell, 1876-1926. New York: Cambridge University Press, 1985.

Samuelson, P., R. Davis, M. Kapor, and J. H. Reichman. "A Manifesto Concerning the Legal Protection of Computer Programs." Columbia Law Review 94, no. 8 (1994): 2308-2431.

Scherer, F. M. "Nordhaus's Theory of Optimal Patent Life: A Geometric Reinterpretation." American Economic Review 62 (June 1972): 422-427.

Scherer, F. M., and D. Ross. Industrial Market Structure and Economic Performance. Boston, Mass.: Houghton Mifflin, 1990.

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

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

United Nations Conference on Trade and Development. The Outcome of the Uruguay Round: An Initial Assessment. New York: United Nations, 1994.

van Dijk, T. W. P. The Limits of Patent Protection. Maastricht, The Netherlands: Universitaire Pers Maastricht, 1994.