Transitions in Agbiotech: Economics of Strategy and Policy

EDITED BY
William H. Lesser

Proceedings of NE-165 Conference
June 24-25, 1999
Washington, D.C.

Including papers presented at the:

International Consortium on Agricultural Biotechnology Research Conference
June 17-19, 1999
Rome Tor Vergata, Italy

PART THREE:
Intellectual Property Rights

16. The Relationship Between Patents and R&D Investment: Biotechnology Patents as Incomplete Contracts

Marc Banik and Jay P. Kesan

Food Marketing Policy Center
Department of Agricultural and Resource Economics
University of Connecticut
and
Department of Resource Economics
University of Massachusetts, Amherst

The Relationship Between Patents and R&D Investment: Biotechnology Patents as Incomplete Contracts

Marc Banik

Jay P. Kesan

College of Law
University of Illinois−Champaign

Chapter 16

The Relationship Between Patents and R&D Investment: Biotechnology Patents as Incomplete Contracts

Marc Banik and Jay P. Kesan[1]

I. Introduction

A. Patents as Contracts: Incentives for Investment in R&D

Biotechnology inventions, such as genetically-modified crops and hormones, synthesized through recombinant DNA technology are the products of large investments in research and development (R&D). By conferring rights to the inventor to exclude others from making, using or selling a patented invention (without a license), patents provide incentives for investment in costly and risky R&D. Patents also encourage the dissemination of information on new inventions, so as to allow competitors to build upon or develop improved versions of patented inventions. While patents may create incentives for investment in R&D, they also impose social costs in the form of reduced levels of competition or wasteful design-around efforts by competitors. Efficient patent systems, therefore, aim to induce investment in R&D, while limiting losses due to market power.

A view that has largely influenced patent enforcement policy holds that market power can be mitigated by limiting the scope of patent protection. This can be accomplished by granting narrowly defined patents as opposed to broad ones, and by maintaining post-issuance opportunities to invalidate patents that fail to meet the statutory requirements of patentability such as novelty, utility and nonobviousness. While these strategies increase public welfare by creating more competitive markets, they also create uncertain property rights which may make it more difficult for patent holders to appropriate the surplus of their inventions. Weak intellectual property rights may thus have an effect of reducing incentives for investment in high technology R&D.

Patent policy makers must therefore consider not only the distortions to perfect competition, but also a policy's distortions to incentives for investment in R&D. In this paper, we propose that viewing patents as *incomplete* contracts is a useful means to explain the relationship between institutions of patent enforcement and investment in R&D. To see why, consider first, why many view a patent as a contract: in exchange for disclosure of the invention, the public's agent, the U.S. Patent & Trademark Office (PTO), grants the patentee a limited right to exclude others from making selling or using the invention. Thus, the *quid pro quo* of the patent system provides the patentee with a limited property right in exchange for an adequate disclosure about the invention. Describing patents as contracts also suggests that the Coase theorem may be applied to propose an optimal patent policy. Patent policy that creates well-defined property rights, leads to efficient use of social resources, such that new innovations will be introduced so as to meet society's demand for new technology. While this is attractive because it would mean that patent policy could be analyzed in the same way other forms of property rights are analyzed, patents differ from ordinary property rights in a number of ways.

Patents, unlike other types of property rights, may be rescinded after they have been granted, either as a result of the patent being invalidated or the scope of patent protection being modified in post-issuance litigation. Thus, a patent may be viewed as a contingent property right. Investors of R&D projects will view the patent in terms of a probabilistic property right, where the probability of invalidation reduces the expected return of an R&D investment project. Another important distinction has to do with the R&D process itself: an investor must commit to R&D expenditure without knowing whether the benefits of the project will accrue to the patentee or not (because the patent may be invalidated). Such information may be costly to obtain prior to investing because not enough is known about the invention until it is made. The public cannot, therefore, "contract" with the inventor to create a new invention, but instead must establish patent enforcement rules by which a patent may be invalidated. Such rules may indeed be difficult to specify in sufficient detail so as to work in the same manner as a property right, as Coase envisioned.

We propose that patents are incomplete contracts because they create contingent property rights, which in turn reduce incentives for investment in R&D. Reduced levels of R&D investment, retard the rate at which new technologies are delivered to the public. The literature on hold-up (Klein *et al*. 1978, Williamson 1979) has explained reductions in investment in specific capital in terms of opportunistic behavior, or informational asymmetry. We wish to distinguish these losses (although possible in the case of patents, as we shall see below) from the losses that occur as a result of the gap-filling of incomplete contracts between the patentee and the PTO. The gap-filling rules or default rules that we are referring to correspond to the substantive rules in patent law (*e.g.*, Section 102 in Title 35 of the U.S. Code) that allow opportunities for invalidating a patent. In other words, these default rules confer broad residual rights to the public to invalidate a patent through post-issuance litigation.

Our theoretical model, applies and builds on Grossman and Hart's (1986) model of incomplete contracts to the case the invention and patenting process. We view the inventor and PTO as vertical partners in a process of technology commercialization that spans steps of R&D investment to patent issuance. The decision variables in such a process are the amount of prior art disclosed during patent prosecution and the presumption of validity accorded to the prior art disclosed by the patentee. The model we propose suggests that in the case of high technology patents, it may be optimal for the PTO to provide incentives to the patentee to produce a complete prior art disclosure, by according a high presumption of validity to the disclosed prior art, which limits the use of the disclosed prior art for invalidation purposes in subsequent litigation.

We begin by defining incomplete contracts and then introduce the technology commercialization process in Section II. In Section III we define the model and in Section IV, we apply it to analyze the impact of different patent policies in high and low technology fields of invention. Conclusions appear in the final section.

B. Incomplete Contracts Defined

Economics and legal scholars have used the term "incomplete contract" differently. The legal appreciation for incomplete contracts has been in understanding how the rights and obligations of each party remain unspecified as a result gaps in the literal language of a contract. A contract is "obligationally" (Ayres and Gertner 1992) or literally incomplete if some of the details of the contract, such as price, delivery time, technical specifications remain unspecified for a set of circumstances in which the contract is to apply. In this regard, much of the legal focus on incomplete contracts is in determining the gap-filling role courts should play in specifying the default rules (such as good faith rules) that apply to contracts.

Default rules eliminate literal incompleteness by specifying the terms of a contract that shall apply in the absence of their literal specification. Any contract is thus made literally complete as a result of default rules that attach to a contract that is literally incomplete. Default rules may take the form of general terms that specify price, delivery date and other details *outside* specific terms of the contract.

On the other hand, the economist's definition of incomplete contracts refers to the ex post efficiency of contractual outcomes. Does the contract allow the joint surplus of the parties to be maximized by taking into consideration the buyer's marginal valuation of the good and the seller's cost? A contract is economically incomplete if it fails to induce Pareto-improving trade in all of the relevant contractual contingencies. Such contracts are thus "state contingent" incomplete (Ayres and Gertner 1992) because the immutable terms of the contract prevent parties from engaging in mutually beneficial trade. Economic incompleteness is viewed with reference to "states of the world" because it is these factors that determine buyer willingness to pay and seller costs. For instance, the value of a gallon of water will be higher to the buyer during a drought than during a flood. Yet, if a contract fixes the price of water (either through literal specification or by default rules) at a price that does not vary with the weather, this market will not clear, leaving some buyers unserved or creating a glut in all but a very small range of prices. These outcomes are not ex post efficient, because both parties could be made better off (buyer could get water, seller could get a better price) if the contract terms were to be renegotiated. The contract is therefore economically incomplete because it does not lead to an ex post efficient outcome.

It is only relatively recently that legal audiences have considered this form of incompleteness, that is the efficiency considerations of trading rules, in deciding the role the courts should play in gap-filling or the interpretation of default rules. Hadfield (1994) analyzes, for example, the roles efficiency-minded courts should play in determining the damages for breach when they are limitedly competent. She proposes that even a limitedly informed court can enhance welfare by enforcing a liability standard, as opposed to a bright line rule, because the former induces changes in the contracting partner's ex ante behavior, while the latter does not. In a similar vein, Ayres and Gertner (1992) have argued, against the conventional wisdom that the gap filling role the courts should assume is to enforce terms that maximize the joint surplus[2] of parties (Goetz and Scott 1981). They contend that courts should maintain penalty defaults against the more informed party to a contract, so as to induce the parties to reveal information that would lead to ex post efficient contracts. These studies emphasize foremost that the terms of the contract itself can have important effects on ex post efficiency as a result of the parties' means of dealing with literal incompleteness.

The relationship between literal incompleteness and economic incompleteness, therefore, depends on the role default rules play in filling the gaps of incomplete contracts. Contracts that avail themselves to many forms of literal incompleteness leave many gaps to be filled by default rules. To the extent that the recourse to default rules create inefficient outcomes (because thin markets are created), literal incompleteness results in economic incompleteness. Economic incompleteness may, however, result due the court's inability to verify whether certain actions have been performed as desired by the contracting parties (*e.g.*, whether all relevant prior art known to the patentee has been disclosed during the course of patent prosecution to the PTO).

Non-verifiability is the premise of Grossman and Hart's (1986) incomplete contracts theory of ownership. In this model, a literally incomplete contract results due to the fact that some activities of the contract may be observable to the person undertaking them, yet non-verifiable to the courts. These activities are what they term "non contractibles." Gaps of literal incompleteness are filled not by default rules, but rather "residual rights" that enable one party to essentially specify the default rules that apply in contingencies outside the literal terms of the contract. Problems of verifiability are typically observed in the case of a two-stage production process, where optimal production decisions of the second stage are contingent upon the outcome of the first stage. A contract between the upstream and the downstream partner may be economically incomplete if it fails to induce the upstream (stage-one) partner from making investments that maximize the joint surplus of the parties. If the contributions of the upstream partner are observable but not verifiable to the courts, (because, for instance, the technology is not separable, or the task to be performed is not programmable), the upstream partner will not have any basis to appropriate the benefits from diligent performance of his duties. This is viewed in the principal-agent model as not being incentive compatible. However, here the important distinction is that, the upstream partner, anticipating that the benefits will be appropriated by the downstream partner, will be reluctant to make surplus-enhancing investments in the first place.

Grossman and Hart suggest that the economic incompleteness of such outcomes can be reduced through an allocation of residual rights to the upstream partner. In the following section, we build on this approach of optimal ownership, by proposing that the

R&D and patenting process can also be viewed as a vertical relationship between the patentee (inventor) and the PTO (public).

C. Patents as Incomplete Contracts

The R&D and patenting process includes many of the sources of economic incompleteness described above. A patent may be viewed as a literally incomplete contract whose gaps are filled by rules of patent enforcement. These rules apply to prosecution, infringement and invalidation. Patents may be viewed as economically incomplete contracts because the prosecution may not make full use of the patentee's information about, for example, the novelty of the invention, because it does not provide adequate incentives for the patentee to reveal such information.

The R&D and patenting process may also be considered a joint production problem, where the patentee's investment in R&D and resulting claims of novelty are non-verifiable. One important distinction is, however, that the R&D investments a patentee makes, do not necessarily improve the surplus (as in the case of the relationship-specific capital assumed in the models of hold-up) to be shared by the patentee and the public. Such investments merely increase the chances of producing that surplus, since the outcomes of R&D projects are not deterministic.

Although the formal model we develop in this paper relies substantially on this latter view of R&D and the patenting process as a joint production problem, the notions of default rules and asymmetric information have clear applications to the patenting process. We discuss these in further detail in Sections II and III.

A patent is literally incomplete, because what qualifies as invalidating prior art under the statutory bars specified in Section 102 of the patent act is subject to judicial discretion. Section 102 of Title 35 of the United States Code requires that an invention be new in order for it to be patentable. If the invention has been described in a printed publication anywhere in the world 12 months prior to the filing date of a patent, the patent may be rendered invalid in post-issuance litigation. Courts have ruled that a printed publication must be accessible to the public in order to qualify as invalidating prior art under Section 102. However, the meaning of a "printed publication accessible to the public" is subject to judicial discretion. For example, in *In Re Hall*, 781 F.2d 897, 228 U.S.P.Q. 453 (Fed. Cir. 1986), the courts ruled that a doctoral dissertation that was catalogued in library archives constituted a printed publication within the grasp of the public's knowledge, despite the fact that it may not been actually accessible to an inventor or scientist.

A patent may be viewed as a incomplete contract because the courts can attach a different legal significance to the claims of a patent *after* it has been issued. In *Amgen, Inc. v. Chugai Pharmaceutical*, 927 F.2d 1200, 18 U.S.P.Q.2d 1016 (Fed. Cir. 1991), the court invalidated Amgen's claim over any DNA segment that would encode for a hormone (erythropoietin) that stimulates red blood cell production, on the grounds that the specification did not enable the broad scope of patent protection sought by Amgen. In *Enzo Biochem v. Calgene*, 14 F.Supp.2d 536 (D. Del. 1998), the plaintiff asserted that Calgene's FLAVR SAVR tomatoes incorporated genetic antisense technology that was covered in Enzo's patents. However, the court held that Enzo's claims of infringement were not substantiated because the Enzo patent was limited to the use of antisense technology in E. coli, since it did not provide enabling information for use in plants and animals.

Outcomes such as these can be viewed as cases where institutions of patent enforcement create economically incomplete contracts because literal incompleteness causes the court to resort to default rules which allow a patent to be invalidated by several different contingencies not anticipated during patent prosecution. In the hypothetical situation of complete contracting, there would have been no uncertainty as to whether a prior art reference would render a patent invalid. Reducing the extent of incompleteness in the patent contract, would thus allow greater investment in R&D projects without incurring risks of not being able to appropriate the benefits of the project. In contrast, greater incompleteness in the patent contract creates opportunities for this type of ex post hold up, reducing the ex ante incentives for investment in R&D.

A patent may be invalidated or the scope of a claim reduced as a result of imperfect information about the state of the art. In our model, we suppose that in the case of high technology fields of invention the patentee and the PTO are asymmetrically informed. Concealment of prior art results in inefficient contracting between the patentee and the PTO. This is because the PTO, not being aware of such prior art, may grant overbroad claims for inventions already within the grasp of the public. This information asymmetry thus allows the patentee to collect an information rent, in the form of a license fee for a patent that is not valid.

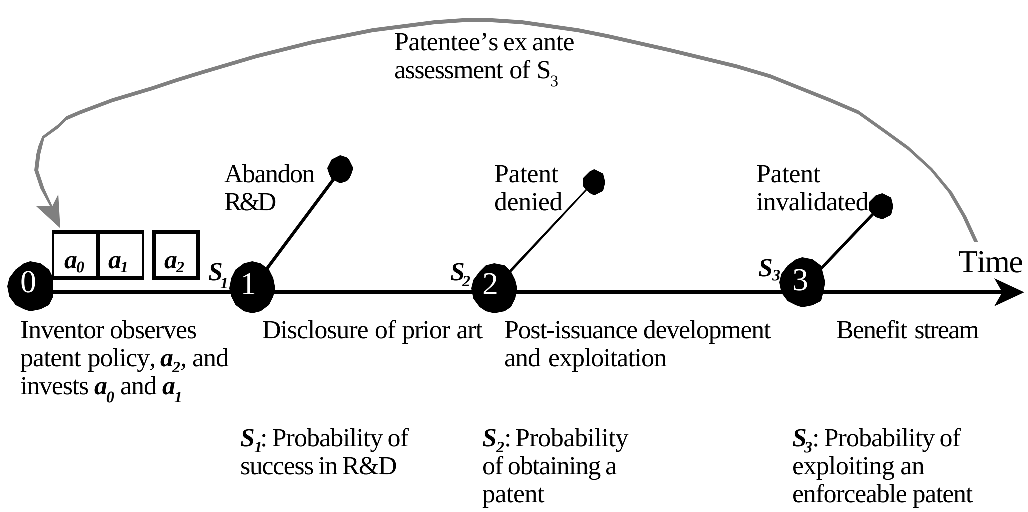
Litigation costs to invalidate such unwarranted patents also exacerbate welfare losses due to imperfect or asymmetric information. A patentee holding a patent that could be invalidated in post-issuance litigation may decide to set a license fee that is lower than the cost of litigation. A potential infringer, not wishing to bear the risk and costs of litigation may pay the license fee for an unenforceable patent. So long as the costs of litigation and information acquisition for other infringers are significant, a patentee may continue to collect several small license fees. Indeed, as Lanjouw and Schankerman (1999) have shown, one of the determinants of the probability that a patent will be challenged is the size of the stakes – which, in this case, corresponds to the license fee. High license fees will make the costs of litigation and information search worthwhile for potential infringers. Low license fees, in contrast, allow the patentee to escape invalidation by discouraging attempts to invalidate a patent.

Considerations like these form the basis of a firm's decision regarding whether to invest in R&D. Models of expected utility maximization, however, do not capture the important effects of public policies such as rules of patent enforcement. Investing in an R&D project is different from an optimal lottery strategy because institutions of patent policy may allow the inventor to have control over the probability of exploiting a valid patent. The control we are referring to here is the information the patentee can provide to the PTO during prosecution. Thus we are concerned with how such policies might affect the decision to invest in R&D. We describe this in the context of the overall R&D and patenting process which we term the technology commercialization process.

II. Patents and the Technology Commercialization Process

A new invention begins as an investment project, the expected value of which depends on the outcome of initial R&D efforts, the probability of obtaining valuable property rights protecting the invention (a patent), and the probability of successfully exploiting that property right (market success). This sequence is described in Fig. 1. At time 0 the inventor evaluates the alternatives for investment in R&D projects, based on their expected net profits, given the public's specification of the terms of the patent regime, a_2 (*e.g.*, duration, eligible subject matter for patent protection) and makes an R&D investment, a_1. If the results of R&D are successful (probability S_1), the inventor applies for a patent and obtains one with a probability S_2 at time 2. Once a patent is obtained, the invention creates an expected revenue stream of monopoly rents or licensing fees for the duration of the patent. However, the validity, enforceability and scope of protection of a patent may be challenged in court [with probability ($1-S_3$)] after it has been issued on a variety of grounds (*e.g.*, inability to meet the requirements for patentability) if such evidence is presented to the court in post-issuance litigation. Two factors that affect the validity of a patent, or its ability to generate benefit streams, are the disclosure of prior art during patent prosecution, and the presumption of validity accorded disclosed prior art. Each of these may be considered decision variables in the technology commercialization process illustrated below.

FIGURE 1 Schematic Representation of the Technology Commercialization Process



A. Disclosure of Prior Art

The prior art is the basis for determining whether the invention sought to be patented meets statutory requirements of novelty and nonobviousness. For example, if a prior art reference, such as a scientific publication, or a printed document, demonstrates that the invention has been used anywhere in the world 12 months prior to the patentee's filing date, the invention is statutorily barred from being patented. Prior art disclosure provides benefits for both the public and the PTO, albeit for different reasons. A complete prior art disclosure reduces the probability (S_3) that a patent will be invalidated in post-issuance litigation since it allows the PTO to evaluate a larger set of prior art references that could be used to later invalidate the patent. A patent granted based on a more complete disclosure of prior art is less likely to be invalidated, because it reduces the number of references that will have not been previously considered by the PTO.

The public benefits of a complete prior art disclosure are that it allows the PTO to grant claims of proper scope that are commensurate with the extent of innovation, because it is better informed regarding the state of the art. This can be particularly important in high technology sectors where the PTO may not have access to all of the prior art that is available to the inventor.[3] The PTO, not being an expert in the field and having limited resources, may not be able to access all the relevant prior art compared to the inventor, who is typically an expert in the field. This is especially true when the relevant prior art is non-patent art such as scientific publications or when relevant prior art information is already in the public domain, as in the case of computer software. In the context of the *quid pro quo*, we may view this as creating information asymmetry between the trading parties, with the PTO being the lesser informed of the two. In contrast, in the case of inventions in low technology or established technology areas, such as the mechanical arts, the PTO is better able to ascertain the state of the art. We therefore assume that the PTO and the patentee would be symmetrically informed in the case of low technology or established technology inventions. These assumptions have implications on the nature of the exchange between the PTO and the patentee. This will become clearer, when we consider the presumption of validity accorded to the disclosed prior art.

B. Presumption of Validity Accorded to Disclosed Prior Art

A patent will be difficult to invalidate if an extensive prior art disclosure is made by the patentee at the time of patent prosecution and that prior art cannot be the basis for patent invalidation by the courts in post-issuance litigation. While a more complete prior art disclosure decreases the probability of invalidation, a high presumption of validity also reduces the probability of invalidation because it restricts the opportunities for the courts in post-issuance litigation to attach a different legal significance to the prior art. For the patentee, a high presumption of validity accorded to disclosed prior art reduces the chances that a patent will be invalidated as a result of the disclosed prior art taking on different legal significance. Under a regime of low presumption of validity in contrast, prior art disclosed during prosecution *can* be reexamined by the courts in a post-issuance lawsuit (*e.g.,* invalidation or infringement suit) in order to render the patent invalid.

Because it removes the possibility to reduce market power, the public may wish to accord a high presumption of validity only in circumstances where the resulting benefits outweigh distortions to competition and reductions in incentives for R&D investment. One such advantage is the more certain property right a high presumption of validity creates. This in turn allows more risky R&D projects to be undertaken (because the expected value of the project is not further reduced by uncertainty over patent rights). This increase in investment in R&D in such high technology projects can outweigh the social losses of market power, because high technology inventions often create significant positive externalities that spill over in other markets. In addition, under a regime that accords a high presumption of validity for disclosed prior art, the PTO may capitalize on a complete prior art disclosure and grant patent rights that are commensurate with the new information in the patent disclosure and avoid the negative consequences brought about by an overbroad grant of patent rights.

C. Characterizing Optimal Patent Policies

Optimal patent policy must weigh, among other things, the inducements provided for investment in R&D against the social costs imposed by patents. The main social cost we are concerned with in this paper is diminished ability to develop new inventions as a result of policy that discourages investment in R&D. Wrongly or improperly issued patents create several difficulties for inventors, which we may view as added social costs. These include the wasteful design-around activities of inventors who try to avoid infringement. The cost of developing new inventions that incorporate patented inventions increases if downstream inventors must pay upstream inventors license fees (Heller and Eisenberg 1998). Existing patents may also raise the financing costs for R&D projects because they increase the possibility that the new invention will infringe upon existing patents. Financiers view this as raising the risk (or lowering the expected return) of the investment project, and thus command a higher risk premium on loans to inventors.

We may view these as the costs of economically incomplete contracting: a patentee may have access to information relevant to patentability that the PTO may benefit from, but may not do so for lack of adequate incentives. One means of correcting this is through a policy that accords a high presumption of validity to disclosed prior art. This would allow the PTO to exploit the patentee's private information by then granting claims of proper scope. In high technology fields of invention, such as biotechnology and computer software, the problem of asymmetric information is particularly acute for reasons mentioned above in Section A. These may be viewed as consequences of imperfect or asymmetric information between the patentee and the PTO.

To the extent that increased information allows proper patents of proper scope to be granted and it reduces the probability of invalidation for the patentee, it reduces economic incompleteness. This is because both the public and the patentee are made better off through a "contract" that allows a high presumption of validity to attach to disclosed prior art. We view this as a case where the objectives of the PTO and the patentee are congruent and model this disclosure and patenting process as a joint production process or a vertical relationship.

Such an analysis does not however account for the option value of being able to retain full rights to invalidate a patent in post-issuance litigation. Because it may be difficult to ascertain what a properly defined high technology patent should encompass, the option of being able to easily invalidate a wrongly-issued patent, may not negligible. If the PTO finds that a basic patent, on a gene sequence for example, should never have been issued, the value of being able to invalidate it, is that it removes what could be a development obstacle for future inventions. For the case of high technology inventions, this option value can be high, if wrongly-issued patents are significant obstacles to the development of important inventions. The public welfare benefits of such an option value are difficult to determine and most likely vary considerably from across different fields of invention. Accounting for this option value in a model of incomplete contracts with asymmetric information is a complex issue since it requires introducing additional assumptions about the information levels of each party and the value of the information through time. Our model does not attempt to do this. As a result we assume that in according a particular presumption of validity to disclosed prior art, the PTO is conscious of the option value it gives up in exchange for information from the patentee and is capable of appreciating the import of the prior art disclosed by the patentee.

III. Modeling Patents as Incomplete Contracts

Consider the three-period model described in Figure 1, in which the inventor (patentee) invests a_0 in R&D at time 0. The patentee also invests a_1 prior to observing the outcome of R&D, in order to assess the intellectual property situation in the field in which R&D activities are contemplated. Investments a_1 allow the patentee to conduct a complete search of prior art in order to determine if the R&D project could lead to an infringing product but also to be able to support claims of novelty and nonobviousness in a patent application. Since the decision to invest in R&D depends on the expected profits, a_0 is a function of a prior belief about S_3, the probability of holding a valid patent, in the post-issuance stage. This depends on the regimes of patent enforcement employed by the courts. Hence, we write $a_0=a_0(S_3)$. In addition, at time 0, the PTO invests a_2, in bureaucratic infrastructure and commits to particular institutions and practices of patent examination.

The results of investments a_0 in R&D initiated at time 0 are contingent upon the success rate S_1. If R&D is successful the patentee applies for a patent at time 1. In the course of applying for a patent, the patentee discloses prior art q_1 and makes a claim on territory not within the prior art. If the patent is issued, then the patentee may exploit the patent by producing the invention and selling it (presumably at monopoly prices) or licensing it. If a patent does not issue (probability $(1-S_2)$), the patentee may choose to re-apply with modifications, appeal the PTO's decision or abandon the project. During the patent application process, the claims may be negotiated in light of information made available to the PTO either from the patentee, or through its own search. Granting a patent, however, does not ensure that the property right is fixed; it may be invalidated in post-issuance litigation, depending on the presumption of validity, q_2, that the public (PTO) has accorded to the cited prior art.

The process in Figure 1 can be thought of as a two-period production model in which each party (the patentee and the PTO) supplies appropriate levels of q_1 and q_2 in order to bring a new invention to the market. The disclosed prior art q_1 is non-contractible in the sense that it cannot be verified by the public since the public cannot determine whether the complete prior art known to the public has been disclosed to the PTO. The presumption of validity accorded to the cited prior art by a court in post-issuance litigation is a patent policy choice that is known at time 0. The non-contractibility of q_1 creates an opportunity to allocate residual rights of control over the variables q_1 and q_2. Hence, we can ask whether joint surplus could be increased by allowing one party to control q_1 and q_2 through the choice of patent policy specified at time 0. With this question, we define the benefit functions of each party.

Rents from successful patenting accrue to the inventor for the duration of the patent, with probability S_3. The net benefits of knowledge disseminated to the public (which includes users, improvers and competitors) as a result of the patent disclosure are captured with certainty. The benefit accruing to each party can be expressed as $Bi[a_0,a_i\boldsymbol{j}_i(q_1,q_2)]$, where i=1 for the inventor and i=2 for the public. We assume that B_i is increasing in $\boldsymbol{j}_i$ for each party. In the case of the patentee $\boldsymbol{j}_1$ corresponds to the probability S_3 of having an enforceable patent. B_1 is thus a realization of expected benefits, given S_3. The PTO similarly wishes to maximize $\boldsymbol{j}_2$, which are the social benefits of patents issued with properly defined scope (*e.g.,* providing enabling disclosure to improvers and competitors in addition to maintaining incentives for R&D).

The problem of the policy maker is, therefore, to decide how rights of control over q_1 and q_2 need to be accorded in order to maximize joint surplus $B_1 + B_2$. To reiterate the problem being modeled, we have defined the following function representing the joint surplus:

$$(1) \qquad B_1[a_0, a_1, \boldsymbol{j}_1(q_1,q_2)]+B_2[a_0, a_2, \boldsymbol{j}_2(q_1,q_2)]$$

where the variables are defined as:

a_0: ex ante investment in R&D by patentee
a_1: ex ante investment in prior art search by patentee
a_2: ex ante investment in patent system (salaries and work-load of patent examiners)
q_1: disclosure of known prior art
q_2: presumption of validity for cited prior art
B_1: patentee benefits (monopoly prices, licensing fees)
B_2: public benefits (knowledge disseminated, consumer benefits)

S_1: probability of success in R&D
S_2: probability of obtaining a patent

S_3: probability of holding an enforceable patent

j_1: patentee benefits as a result of disclosure (q_1) and presumption of validity (q_2).
j_2: public benefits from disclosure (q_1) and presumption of validity (q_2)

For simplicity we assume that $S_3 = j_1$, implying that patentee benefits of control over q_1 and q_2 are increased probability of capturing rent streams. Since the present analysis is concerned only with cases where the patentee patents the product of a successful R&D project, we assume that S_1 and S_2 are independent of control over q_1 and q_2.

A. Specification of the Model Under High and Low Technology Environments

The respective benefit functions of the patentee and the PTO will be different for the case of high technology (*e.g.*, biotechnology or computer software) and low technology or established technology inventions, since the costs and the benefits of operating in these environments differ. These differences are captured in the respective φ_i functions which we define as linear combinations of q_1 and q_2, expressed below:

(2) $j_1 = a_1 q_1 + b_1 q_2 + e_1 C$ (patentee)

(3) $j_2 = a_2 q_1 + b_2 q_2 + e_2 C$ (PTO)

The functions are distinguished by the values for the coefficients a_n, b_n [4] and e_n, which are weights on the relative importance of non-contractibles (q_1) and contractibles (C). The magnitude of the coefficient indicates the relative importance of the factor (hence degree to which control will matter), while the sign of the coefficient indicates whether increases in the value of the variable (q_1 or q_2) result in an increase or decrease in benefits. A negative coefficient indicates that control of the variable by party j adversely affects the benefit function j_i. This is the case of incongruent objectives. The differences in the coefficients for the high and low technology sectors are due to information asymmetries and are explained below.

1. High Technology Sectors. In high technology sectors, R&D techniques themselves evolve rapidly, such that an inventor, being an expert in the field has relatively better knowledge of the state of the art, than the PTO with restricted information resources. Asymmetric information between the patentee and PTO is therefore assumed. The benefits for the PTO in obtaining a more complete disclosure is that it will be able to grant claims of proper scope. The coefficient a_2 will thus be large and positive. Since overly broad claims may prevent other inventors from creating improved products or applying the knowledge disclosed in the patent in new areas, the public values information about the prior art relatively more than its right to invalidate a patent. The coefficient b_2, on the presumption of validity (q_2) accorded to the patent will therefore be small and positive.

Because the costs of and risks of R&D in high technology are high, once a patentee obtains a patent, it must be stable, in the sense of reducing the post-issuance chances for invalidation. This will allow the high costs of R&D to be offset by high expected profits. The weight on q_2, $\boldsymbol{b}_1$, will thus be positive and large. As the patentee discloses more prior art information he increases his chances of obtaining a stable patent. The weight on q_1 will thus depend on the degree to which complete prior art disclosure can reduce the probability of invalidation. If there is a strong relationship, due to an incentive provided by the PTO, $\boldsymbol{a}_1$ will be large and positive. This is one of the policy options examined in Section IV.

2. Low Technology or Well-Established Technology Sectors. In low technology sectors, we assume symmetric information between the patentee and the PTO. Since the relevant prior art can be easily identified by the PTO, it will be most concerned with not giving up its right to invalidate the patent in exchange for information it could obtain from sources other than the patentee. Thus, for parties with symmetric information, disclosure of prior art will be relatively unimportant for both the PTO and the patentee. This is particularly so since the patentee does not receive any incentives to conduct a thorough prior art search and disclose it to the PTO. The coefficients $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ will thus be small and positive indicating that either party would be indifferent to disclosure by the PTO or the patentee. Since low technology inventions involve less cost and risk for the inventor, the PTO would prefer to accord a higher presumption of validity only in cases where it induces investment in risky high technology R&D that create large social benefits. Granting a high presumption of validity therefore reduces $\boldsymbol{j}_2$ by relatively large amounts, for the case of low technology inventions. Thus $\boldsymbol{b}_2$ is large and negative.

Because the probability of success S_1, is generally higher in the low technology case, the expected profits from an R&D project may be higher for a low technology project. The low technology inventor could thus afford to bear a higher probability of invalidation that would result from a policy of a low presumption of validity. The coefficient $\boldsymbol{b}_1$ is thus small and positive. Given that both $\boldsymbol{a}_1$ and β_1 are small, most of the weight will be on the contractibles (C), or factors such as brand names, marketing channel agreements – conditions not relying on patents or R&D. In this case $\boldsymbol{e}_1$ will be large. The assumptions we make on the parameters of each of the $\boldsymbol{j}_i$ functions are summarized in Table 1.

In high technology sectors a more complete disclosure of prior art would be Pareto-improving, since both coefficients $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are positive and large. Since $\boldsymbol{b}_2$ is small, both parties could be made better off by a policy in which presumption of validity is "traded" for a more complete disclosure of prior art, supplied by the relatively more informed patentee. The objectives of the patentee and the PTO are thus congruent. For low technology inventions, the objectives will be incongruent because $\boldsymbol{b}_2$ is a large negative coefficient, indicating granting stable property rights on low risk R&D would penalize the public's benefit function.

Table 1

Coefficient	Description	High Technology		Low Technology	
		Magnitude	Sign of Coeff.	Magnitude	Sign of Coeff.
α_1	Relative importance of prior art for patentee (*i*=1) and PT0 (*i*=2)	Large	+	Small	+
α_2		Large	+	Small	+
β_1	Relative importance of presumption of validity	Large	+	Small	+
β_2		Small	-	Large	-
ε_1	Relative importance of contractibles	Insignificant	+/-	Large	+/-
ε_1		Insignificant	+/-	Large	+/-

The tradeoff that the PTO must consider is how increases in the presumption of validity decrease public benefits B_2, by reducing the opportunities for invalidation on the one hand, and at the same time reduce the incentives for investment in R&D on the other hand. These effects are described in equation (4) below. For low technology inventions, we have the following relationship:

(4)
$$\frac{\partial B_2}{\partial \varphi_2} > \frac{\partial B_2}{\partial \alpha_0}_{LowTech}$$

This is because the productivity benefits of low technology R&D are small in relation to the welfare benefits of reducing market power. Since low technology[5] R&D produces only small shifts of production frontiers, the welfare benefits of reducing market power by maintaining a low presumption of validity[6] offset any benefits that would result from increased investment in low technology R&D. In contrast, for high technology R&D, we have the following relationship:

(5)
$$\frac{\partial B_2}{\partial \varphi_2} < \frac{\partial B_2}{\partial \alpha_0}_{HighTech}$$

The effect of increases in R&D can thus outweigh the social losses due to market power that occur as a result of a high presumption of validity that is granted to disclosed prior art. This is because high technology patents disclose revolutionary teachings or methods which the public may build upon or use. The public welfare losses from having to pay license fees in this case, are believed to be relatively small in comparison to the benefits the new technology provides. It should be clear that equations (*4*) and (*5*) are assumptions, and not results of our model. These are consistent with the parameter values of Table 1 in which it is assumed that each agent maximizes their respective φ_i functions. It is based on these assumptions that we propose optimal patent enforcement policies for both high and low technologies.

IV. Analyzing Alternative Patent Policies

We consider two policies, which are distinguished by the incentives provided by the PTO to the patentee to disclose all relevant prior art. Policy 1 "trades" a high presumption of validity for information on prior art that may not be available to the PTO. Under this policy, prior art disclosed at the time of patent prosecution cannot be used in subsequent litigation to invalidate the patent, except in very limited circumstances.[7] The patentee can thus effectively reduce the opportunities for invalidation by citing all relevant prior art to the PTO. Since a high presumption of validity attaches to only the prior art disclosed by the patentee, we write $q_2=q_2(q_1)$. The benefit functions under Policy 1 are thus:

$$(6) \qquad \boldsymbol{j}_1 = \boldsymbol{a}_1 q_1 + \boldsymbol{b}_1 q_2(q_1) + \boldsymbol{e}_1 C$$

$$(7) \qquad \boldsymbol{j}_1 = \boldsymbol{a}_2 q_1 + \boldsymbol{b}_2 q_2(q_1) + \boldsymbol{e}_1 C$$

Under Policy 1, the patentee discloses the relevant prior art in relation to the desired presumption of validity, which we assume to be high. We denote q_1 and q_2 under this policy as $(q_1,q_2)=(q\text{¢}_1, q\text{¢}_2)$. Under Policy 2, there is no such exchange between the patentee and the PTO. In this case the PTO conducts its own prior art search, and information known only to the patentee is not accorded any special legal significance. In other words, the PTO maintains no presumption of validity with respect to the disclosed prior art. The form of the $\boldsymbol{j}_i$ functions is exactly as initially specified, and are reproduced below as:

$$(8) \qquad \boldsymbol{j}_1 = \boldsymbol{a}_1 q_1 + \boldsymbol{b}_1 q_2 + \boldsymbol{e}_1 C$$

$$(9) \qquad \boldsymbol{j}_2 = \boldsymbol{a}_2 q_1 + \boldsymbol{b}_2 q_2 + \boldsymbol{e}_2 C$$

We denote the disclosure of prior art and presumption of validity as $(q_1,q_2)=(q^2_1,q^2_2)$. The effects of each policy in both high and low technology scenarios are examined below.

A. Policy 1: High Technology

Given the relative importance of the high presumption of validity ($\boldsymbol{b}_1$ is large), the patentee maximizes $\boldsymbol{j}_1$ by disclosing as much as information as possible in order to capitalize on the high presumption of validity. The patentee's disclosure and the high presumption of validity ($q\text{¢}_1$, $q\text{¢}_2$) thus maximizes $\boldsymbol{j}_1$, but will also approximately maximize $\boldsymbol{j}_2$, because both $\boldsymbol{b}_2$ and $\boldsymbol{j}_2$ are small. The result occurs because the patentee is made better off by a policy in which a presumption of validity attaches to all relevant prior art disclosed to the PTO. The PTO is also able to accord patent claims of proper scope, in light of the additional information it receives (that it might not have otherwise) from the patentee. Policy 1 applied to the case of high technology thus leads to higher joint surplus, due to coordination between the PTO and the patentee. Since $\boldsymbol{j}_1$ is increasing in $q\text{¢}_1$ and $q\text{¢}_2$ such coordination will lead to higher levels of ex ante investment in R&D by the patentee which increases both B_1 and B_2. This is because increasing the φ_1 function, corresponds to increasing the expected value of the investment project, because $S_3 = S_3(\varphi_1)$.

B. Policy 1: Low or Well-Established Technology

Policy 1 applied to the case of low technology does not maximize public benefits B_2 because the incentives provided by the PTO are "wasted" in that they do not lead to much more information in order to offset the losses in φ_2 caused by a high presumption of validity. The patentee discloses all prior art, yet the PTO is symmetrically informed (assumption of low technology), which implies that $q'_1 \approx q''_1$. Thus, φ_2 is not maximized, but instead approximately minimized, since β_2 is large and negative in equation (7). The increase in φ_1, brought about by a high presumption of validity increases α_0. However, the resulting benefits are small and do not compensate for the loss in presumption of validity.

C. Policy 2: High Technology

Policy 2 which sets a low presumption of validity for prior art removes incentives for the patentee to disclose prior art. Since the PTO conducts its own prior art search, φ_2 is not maximized, because the PTO is necessarily less informed than the patentee. Thus, $q'_1 > q''_1$. Hence φ_1 function is not maximized by q''_1. Furthermore, given $q'_2 > q''_2$ and a small negative coefficient β_2, φ_2 is not maximized. Because the PTO "controls" the presumption of validity $\varphi''_1 < \varphi'_1$, which implies that there will be less investment in a_0 under Policy 2, than under Policy 1. This is because, in our model, the lower level of φ_1 corresponds to a lower probability of having a valid patent (S_3). Under a policy which accords a low presumption of validity, the investor's ex ante assessment of the probability of being invalidated, causes him to assign a lower expected return to the investment project. This makes investment in high risk (high technology) projects less attractive than investment in other projects that have higher private returns, but lower social returns (*e.g.*, low technology projects). Benefits $B_1 + B_2$ are thus not maximized when the PTO relies on its own prior art search in the case of high technology inventions.

D. Policy 2: Low or Well-Established Technology

With no information asymmetry, a policy that maintains a low presumption of validity will approximately maximize φ_2 since the coefficient β_1 is small. Since α_1 and β_1 are both small, any policy affecting these variables directly or indirectly will have a small effect on φ_1. In this case Policy 2 is efficient in the sense that it does not "waste" the privilege of according a high presumption of validity in order to induce R&D in on low technology inventions. Even though this maintains a higher probability of invalidation for the patentee, its effect in reducing incentives for R&D investment is less acute than the case of high technology because a higher probability of technical success

(S_1) compensates for a lower probability of owning a valid patent (S_3). Policy 2 applied to the case of low technology, therefore, approximately maximizes $B_1 + B_2$ because of the public welfare benefits it maintains through a low presumption of validity.

V. Conclusions

We begin with a familiar concept that a patent can be viewed as a *quid pro quo* between the patentee and the public, trading a limited right to exclude for an adequate disclosure of the invention. Because of the legal regimes governing post-issuance patent litigation which allow a patent to be invalidated by the courts after it has been issued, a patent may be considered an economically incomplete contract between the patentee and the public. We use the incomplete contracts framework to understand the welfare tradeoffs of market power and the timely development of new technologies. Our approach in developing a model of patents as incomplete contracts is to ask whether the joint surplus of a partnership can be increased if there is coordination achieved by one party taking control of the other's rights. The answer to this question is not obvious because control of rights or assets reduces the bargaining power over ex post surplus, while it can increase joint benefits if the objectives of both parties are aligned. We apply this model to the technology commercialization process where the patentee and the public (acting through its agent, the PTO) are seen as vertical partners. Our model thus adopts a systems approach by incorporating R&D decision making and public policy from the perspective of each party.

Describing patents as incomplete contracts allows us to consider the efficiency effects of alternative patent policies that allocate control over two variables (prior art and presumption of validity) to one party. We consider two specific patent policies. Under Policy 1, prior art disclosed by the patentee to the PTO during patent prosecution is accorded a strong presumption of validity in post-issuance litigation, limiting the use of cited prior art to invalidate the patent. Under Policy 2, there is no presumption of validity accorded to prior art cited by the patentee.

These two policies are then considered in the context of low and high technology environments, which are distinguished by the costs of invention and the level of information asymmetry between the patentee and the PTO. In high technology sectors, such as biotechnology, the patentee is relatively more informed about the relevant prior art with respect to an invention as compared to the PTO. The high technology sector is further distinguished by the benefits R&D procure to society, as well as increased risks associated with high-tech R&D projects. Given the costs and benefits of innovation in each scenario, we consider which patent policy could best maximize joint surplus.

For the case of high technology inventions, Policy 1 (high presumption of validity accorded to disclosed prior art) induces higher levels of ex ante investment in R&D. The public efficiently trades a strong presumption of validity for information about the prior art. Since prior art is valued by both parties (albeit for different reasons), Policy 1 that encourages its disclosure, could be viewed as a transfer of residual rights that is welfare-enhancing. When the objectives of the patentee and the PTO are not aligned as in the case of low or well-established technologies, Policy 1 would not maximize public welfare since the benefits of increases in investment are outweighed by the costs of reduced opportunities for the public to invalidate the patent.

For the case of low technology inventions, Policy 2 (low or no presumption of validity accorded to disclosed prior art) is optimal, since it allows the public to curtail losses due to market power by maintaining opportunities for invalidating a patent. In low technology inventions, the increased bargaining has only a small effect on reducing incentives for investment, since R&D projects of this type are less expensive and undertaken with limited regard to patent protection. The main reason why these policy prescriptions apply is that it accounts for the positive externalities that different technologies create, while assuming that the benefits from high technology are a result of public policy incentives. The incomplete contracting framework allows us to view patent policy problem of choosing appropriate incentive schemes as a function of informational and technical (productivity) considerations.

The optimal policy for high technology inventions, such as in biotechnology (characterized by asymmetric information and high productivity effects) provides incentives for well-informed patentees to reveal information regarding the relevant prior art to the PTO during patent prosecution. This is consonant with a body of contracting literature (Ayres and Gertner 1992) that proposes that when parties are asymmetrically informed, default rules that penalize the more informed party will be welfare enhancing since the party that is more informed will be induced to reveal information. However, in our case we do not impose a penalty on the better informed party, the patentee. Rather, we permit a transfer of ex post bargaining rights (*i.e.*, reduce the public's residual right to invalidate the patent) in order to induce the better informed party to reduce the informational asymmetry between the patentee and the PTO. In this regard, the objectives of the patentee and the public are mutually aligned since the reduction in ex post bargaining also creates incentives for higher R&D investment.

Our simple model assumes that the costs and benefits of information and different technologies are exogenously determined. In practice they are part of a larger problem in which the policy environment and private strategies evolve together. As new policies are implemented, decision-makers “update” their R&D investment decision criteria. Developing dynamic models of this process would therefore be a worthy research endeavor, but by no means easy – investment planning in high technology increasingly incorporates complex models of resource management. New models must view R&D organizations not only as rational decision-makers, but also as sources of new capabilities, because the inventive process in itself disseminates new knowledge through enabling disclosures in patents. These can be important factors to consider in high technology R&D, in addition to the trade-off between market power and incentives to invest considered in our model of the technology commercialization process.

Endnotes

[1]Marc Banik is a Ph.D. Candidate, Dept. of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. Jay P. Kesan is Assistant Professor of Law, University of Illinois, College of Law, 504 E. Pennsylvania Ave., Champaign, IL 61820. All correspondence may be addressed to kesan@law.uiuc.edu. The financial support of the University of Illinois Campus Research Board is gratefully acknowledged.

[2]In this case the surplus refers to the profits of the parties to the contract, given that investments have already been made. In our model, we consider the case where joint surplus is increased as a result of incentives to invest in surplus-enhancing activities, such as R&D.

[3]In this respect, the PTO held public hearings in June of 1999 on the subject on identifying the sources of information for prior art references and the methods it should use to gain access to prior art in emerging fields, such as computer software and biotechnology.

[4]We define contractibles as those elements of patentee or public (PTO) benefits that are independent of q_1 and q_2. These include profits resulting from brand names or other sources of market power.

[5]By low technology inventions, we have in mind simple mechanical inventions such as a new type of corkscrew. The social benefits conferred by such devices is likely to be smaller than those of a high technology invention such as a cure for cancer or AIDS.

[6]Recall that $\boldsymbol{j}_2$ is maximized by minimizing q_2, since $\boldsymbol{b}_2<0$.

[7]A threshold test for invoking disclosed prior art might be the following: A court will not invalidate a patent based on disclosed prior art unless it is convinced that no reasonable examiner would have allowed the patent in light of the disclosed prior art. In other words, disagreement about what the disclosed prior art teaches would not suffice to invalidate the patent.

References

Ayres, I. and R. Gertner. 1989. Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules. *Yale Law Journal* 99(1):87-130.

Ayres, I. and R. Gertner. 1992. Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules. *Yale Law Journal* 101(4):729-773.

Goetz, C. J. and R. E. Scott. 1981. Principles of Relational Contracts. *Virginia Law Review* 69(6):967-1024.

Grossman, S. J. and O. D. Hart. 1986. The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy* 94(4):691-719.
Hadfield, G. K. 1994. Judicial Competence and the Interpretation of Incomplete Contracts. *Journal of Legal Studies* 23(1):159-184.
Heller, M. and R. Eisenberg. 1998. Can Patents Deter Innovation? The Anticommons in Biomedical Research. *Science,* 28(May):698-701.
Hermalin, B. E. and M. L. Katz. 1993. Judicial Modification of Contracts Between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach. *Journal of Law Economics and Organization* 9(2):231-255.
Klein, B., R. G. Crawford and A. Alchian. 1978. Vertical Integration, Appropriable Rents, and the Competitive Contracting Process. *Journal of Law and Economics* 21(2):297-326.
Lanjouw, J. O. and M. Schankerman. 1999. Patent Suits: Do They Distort Research Incentives? Mimeo, London School of Economics and Yale University.
Williamson, O. E. 1979. Transaction-Cost Economics: The Governance of Contractual Relations. *Journal of Law and Economics* 22(2):233-261.