I Introduction
In early work on optimal patent design Nordhaus (1969) focused on selection of an optimal patent life, chosen to strike an appropriate balance between the need to stimulate research and the desire to avoid monopolistic production of the invention. More recently, the scope (Gilbert and Shapiro (1990) and Klemperer (1990)) and timing (Scotchmer and Green (1990)) of the patent grant have also been studied as instruments of patent policy. I investigate the same trade-off analyzed by Nordhaus but I consider the instruments of patent validity and antitrust policy, and take patent scope, timing, and life as given. The motivation for this choice is that it is more representative of American patent policy than the use of patent life as an instrument, and it allows me to explore the interaction of patent litigation, output restriction, and the incentive to innovate, more easily than if I considered patent scope or timing. In addition, the model developed below captures many of the salient policy issues associated with chemical or pharmaceutical product innovation.

A patent system in which all patents are valid and have a fixed term is a crude method of promoting research. A patent is desirable in cases in which it raises research expenditure in projects where investment would be too small (relative to the social optimum), but it can also lead to excessive research investment. The performance of the patent system could be improved by making the reward to the innovator a continuous choice variable controlled by a regulator. Varying the life of the patent would be one method of achieving this result. An alternative method of making the reward continuous is to link the probability of a patent grant to the nature of the research. Such a system could induce the choice of any level of research expenditure between the levels of expenditure induced by the no-patent and certain-patent alternatives, while producing less expected deadweight loss than the certain patent system. Actually, in the US patent system, uncertainty about the validity of the patent grant enters at the litigation stage rather than the patent approval stage. Hence my analysis is complicated by the possibility of patent litigation and settlement.

The purpose of this chapter is to characterize an optimal patent policy
using the probability of patent validity as the policy instrument. The optimal policy is then compared to the practice of American patent law. In addition, policy issues concerning the intersection of patent and antitrust law are considered.

I study an environment in which inventors of new products sometimes license their patents to avoid a trial on the issue of patent validity. This model stands in contrast to models of cost reducing process inventions in which patentees may license their patents to improve the technology held by their rivals (see Gallini and Winter (1985), and Katz and Shapiro (1985) and (1987)). Both factors are undoubtedly important in explaining why process innovations are licensed, but with a model of new product innovation, it is possible to study licensing motivated solely by the threat of litigation.

Whether trial occurs in equilibrium depends on the magnitude of the joint profit from a patent license. Antitrust law treats patent licenses differentially, allowing restraints in license agreements that tend to restrict output to the monopoly level. But like cartel members, the parties producing the patented product have an incentive to cheat on the terms of the license and raise their output. Inefficiencies associated with contract formation and enforcement depress industry licensing profits. If licensing causes industry profits to fall below the monopoly level, then trial may occur in equilibrium because joint trial costs may be less than the drop in expected joint profit. The litigation or licensing activities have two implications for the efficiency of the patent system: they may reduce the deadweight loss associated with the patent monopoly (but trial imposes a new social cost), and they affect the payoffs to the winner and loser of the research contest, thereby influencing the choice of research expenditure.

My characterization of the optimal probability of validity yields several noteworthy results. First, a strong antitrust policy should be complemented with a strong patent policy (in the sense of a relatively high probability of validity), and conversely, a weak antitrust policy should be complemented with a weak patent policy. Second, I show that an optimal patent policy may lead to trial in equilibrium. This is possible because a high probability of validity (which would render the threat of trial not credible) may cause overinvestment in research, while a small probability of validity (which leads to licensing) may cause underinvestment. Thus, the sacrifice of trial costs may be necessary to get the ‘right’ level of research. Third, patent litigation has greater social value when antitrust policy is weak; this is true because the social value to a patent challenge grows relative to the social value of settlement as settlement grows more effective at imposing output restriction. Patent challengers can be encouraged to challenge weak
patents, for example, by fee-shifting or through policies that give challengers more control over the timing and location of patent lawsuits.

Finally, I compare the nonobviousness standard of US patent law with an optimal patent policy. The nonobviousness standard is legalese for a requirement that the innovative process display some quantum of inventiveness. Roughly speaking, the more surprising or more difficult an innovation was to achieve, the more likely it is to satisfy nonobviousness. In the model below, I interpret the elasticity of the probability of successful innovation with respect to research investment as a measure of the obviousness of the innovative process. I show that the optimal probability of validity is inversely related to obviousness.

What is the intuition for this result? Other things being equal, the optimal probability of validity is low for a research technology such that the probability of success is relatively insensitive to research investment. Such a technology leads to excessive private research because firms race to get a patent. A low probability of validity reduces the rents to the winner of the patent race and thereby diminishes investment. Similarly, the optimal probability of validity is high when the research outcome is relatively sensitive to effort, because private investment tends to be too low relative to the social optimum since the patentee does not appropriate the full social value of the innovation.

In contrast, I show that the value of the innovation is not monotonically related to the optimal probability of validity. To see why, consider two unrelated new product innovations that result from the same kind of research technology, but the second has greater demand than the first. If the probability of validity is the same in both cases, then the socially optimal and the equilibrium levels of research are both higher in the second case. But it is unclear whether the optimal or equilibrium investment rises faster. It may be necessary to raise or lower the probability of validity to maintain equality between the social optimum and the private equilibrium.

II A model of research, patent licensing, and litigation

I study a two-player, five-stage, complete information model of innovation. Briefly, the stages are: (1) simultaneous choice of research investment, (2) a decision whether to infringe a patent, (3) a settlement offer, (4) acceptance or rejection of the offer, and (5) a decision whether to continue to trial after settlement rejection. In stage (1) a pair of identical firms make simultaneous investments in research at a cost of \( x_i \). The probability that firm \( i \) will get a patent is given by \( p_i(x_1, x_2) \). If neither firm is successful, then there is no further action. If one of the firms obtains a patent (the patent system does not allow a patent grant to more than one inventor), then the firms continue to stage 2.
Stages (2) through (5) model patent litigation. At stage (2) the loser of the research contest chooses whether or not to compete in the market created by the patented invention. Entry into the new market requires the loser to make an irreversible investment in production or marketing. This investment makes the loser liable for patent infringement. For convenience I assume the cost of entry is zero. The only viable defense to a patent infringement suit in the model is a showing of patent invalidity due to nonobviousness. The infringer can prove invalidity with a probability $\alpha$. This probability is commonly known.

Given infringement at stage (2) the winner of the research contest makes a settlement demand to the infringer at stage (3). A settlement agreement might specify royalty terms under which the infringer is allowed to continue production. It might also contain other output restrictions. I am not concerned about the details of the settlement agreement and simply suppose that it leaves a profit of $Z$ to the infringer.

At stage (4) the infringer accepts or rejects the settlement demand. The game ends with acceptance. Rejection leads to stage (5). At stage (5) the patentee either brings an infringement suit or drops the case. If the case goes to trial then the patent is upheld with probability $1 - \alpha$ or invalidated with probability $\alpha$. Both parties incur a trial cost of $T$.

After settlement or trial, production takes place in accordance with policies specified by patent law, antitrust law or the licensing contract. Industry profit is given by $V_1$, $V$, or $2V_0$ when production occurs under monopoly, under duopoly with a license, or under duopoly given patent invalidity, respectively. I assume that $V_0 > T$ and $V \in [2V_0, V_1]$ where the magnitude of industry profit under a settlement license depends on two factors. The first is the application of antitrust law to patent licenses. Antitrust law is more tolerant of contract terms that restrict output in the context of patent licenses than in other contractual settings. The range of $V$ accommodates a range of antitrust regimes that differ in permissiveness toward patent licenses.\footnote{Second, even though antitrust law tolerates output restrictions facilitated by patent licenses, the normal temptations to cheat that face any cartel create transaction costs that erode industry profit under a settlement license. The likely result is that duopoly licensing profits are less than monopoly profits. I define $S_0$, $S$, and $S_1$ to be the social value of production of the new product corresponding to industry profit levels $2V_0$, $V$, and $V_1$, where $S_0 \geq S \geq S_1$.}

The subgame perfect Nash equilibrium is found by working backwards from the end of the game. Equilibrium of stages (2) through (5) may result in trial, no entry, or two different types of licensing agreements. In one
type of license the patentee does not have a credible threat of bringing an infringement suit; the license is merely a collusive device for raising industry profit. In the other type of license both parties could get positive expected payoffs from trial, and the license is a bona fide settlement of litigation.

At stage (5) the patentee will not go to trial unless the expected payoff is higher than the alternative of sharing the market with the infringing entrant. Since the patent is invalidated with probability $a$, trial has a higher expected payoff to the patentee if:

$$(1 - \alpha)V_1 + \alpha V - T \geq V_0$$  \hspace{1cm} (11.1)$$

If condition (1) fails then the loser of the patent contest would infringe at stage (2). The patentee would offer a license giving a profit of $Z = V_0$ to the infringer. The purpose of the license is not to avoid trial, but to use the ‘ancillary’ output restrictions available in a patent license to raise industry profit. The infringer would accept the license and the profit to the patentee would be $W^C = V - V_0$, the profit to the infringer would be $L^C = V_0$, and the social value of the innovation gross of research expenditures would be $\Sigma^C = S$. (The $C$ indicates collusive settlement; $W$ indicates winner of the patent contest; $L$ indicate loser; and $\Sigma^C$ represents the gross social value of the innovation accounting for trial outcomes and costs.)

Now suppose that the patentee has a credible threat of going to trial, i.e., condition (1) is satisfied. The next issue is whether the loser of the patent contest has a credible threat of going to trial. This is equivalent to condition (2) which states that an infringer gets a non-negative expected profit from trial.

$$\alpha V_0 - T \geq 0$$  \hspace{1cm} (11.2)$$

If this condition is not satisfied, then the loser of the research contest will not enter at stage 2. The result is a profit to the patentee of $W^N = V_1$ a profit to its competitor of $L^N = 0$, and a social value of the innovation of $\Sigma^N = S_1$. (The $N$ represents no entry.) When conditions (1) and (2) are satisfied entry occurs and settlement emerges as the bargaining outcome when it maximizes joint profit, thus the condition:

$$V + 2T \geq 2\alpha V_0 + (1 - \alpha)V_1$$  \hspace{1cm} (11.3)$$

must be satisfied for settlement. The payoffs to the patentee and its rival and the social value in the case of a settlement license are:
\[ W^S = V - \alpha V_0 + T, \]
\[ L^S = \alpha V_0 - T, \]
\[ \Sigma^S = S \]  
\[ (11.4) \]

To distinguish this case from collusive settlement I will call it regular settlement. Finally, when both conditions (1) and (2) are satisfied, but condition (3) fails to hold, the parties go to trial to test patent validity. The payoffs are:

\[ W^T = (1 - \alpha) V_1 + \alpha V_0 - T, \]
\[ L^T = \alpha V_0 - T, \]
\[ \Sigma^T = \alpha S_0 + (1 - \alpha) S_1 - 2T \]  
\[ (11.5) \]

Since the parameter \( \alpha \) which represents the probability of patent invalidity is central to the normative analysis below, it is worthwhile to note the relationship between \( \alpha \) and the various litigation outcomes. For patents with a small probability of invalidity condition (2) fails and the loser of the patent race does not enter. For large values of \( \alpha \) condition (1) fails and the parties agree to a collusive patent license. For intermediate values of \( \alpha \) both parties have a credible threat and either trial or settlement may occur. Inspection of conditions (1), (2), and (3) shows that trial does not occur for any value of \( \alpha \) unless the following condition is satisfied:

\[ V_1 T < V_0 (V_1 - V) \]  
\[ (11.6) \]

Thus, if the cost of trial is sufficiently low, then intermediate values of \( \alpha \) may lead to trial. The rival is willing to go to trial because \( \alpha \) is large enough to make the expected value of trial larger than the cost. The patentee favors trial over settlement because licensing causes a dissipation of the rents attributable to the patent \((V_1 - V)\). Notice, in particular, that trial never occurs if \( V = V_1^c \). \(^8\)

Inspection of conditions (1), (2), and (3) also shows that settlement does not occur for any value of \( \alpha \) unless the following condition is satisfied:

\[ V_1 T \leq V_0 (V_1 - V_0) \]  
\[ (11.7) \]

I assume that (7) always holds.
B Investment in research

The first stage of the model incorporates the familiar contest to win a new product patent. Two firms simultaneously choose $x_i$, a research expenditure; if no firm successfully invents a new product there is no further action; if some firm is successful it receives a patent; only one patent is granted. The firms make their investment decisions knowing that the government’s patent policy implies a certain probability of invalidity $\alpha$, and foreseeing the equilibrium outcome of the litigation process, and the implied payoffs to the winner and loser of the patent contest.

The expected profit of research to firm $i$ is:

$$\pi_i = p_i(x_i, x_j)W + p_j(x_i, x_j)L - x_i, j \neq i$$

This expression states that firm $i$ wins the patent contest with probability $p_i$ and earns the expected payoff $W$. Firm $j$ wins with probability $p_j$ and firm earns $L$. Firm $i$ incurs the cost $x_i$ from its research investment. The precise values of $W$ and $L$ are determined by the choice of $\alpha$ and which equilibrium outcome arises in stages (2) through (5).

Restrictions are imposed on the functions $p_1$ and $p_2$ such that the research technology leads to a unique symmetric Nash equilibrium. These conditions are given in the Appendix to this chapter. Let $x = f(W, L)$ be the equilibrium expenditure by each firm as a function of the payoffs to the winner and loser of the patent contest. Natural restrictions on the technology (which are also given in the Appendix) imply:

$$\frac{\partial f}{\partial W} > 0 > \frac{\partial f}{\partial L}$$

So the equilibrium investment increases in the payoff to the winner and decreases in the payoff to the loser.

The amount of research investment stimulated by a patent contest may be greater or less than the socially optimal amount. Problems arise because the private and social value of innovation generally diverge, and because competitors race to be the first to complete an innovation in order to secure the patent. The result can be investment that is either too small or too large (compared to the social optimum), usually accompanied by inefficient duplication of research efforts.

III Optimal patent policy

The optimal patent policy controls the equilibrium research effort indirectly by fixing the probability of invalidity $\alpha$, which determines the contest payoffs $W$ and $L$, which in turn determine equilibrium investment $x^*$. The probability of invalidity also determines the social value of the innovation.
gross of research costs. Thus, the problem of the social planner is to search the feasible set of \((x^*, \Sigma)\) pairs to maximize expected social welfare.

The social planner’s problem is to choose \(\alpha \in [0,1]\) to maximize expected social welfare, which is given by:

\[
ESW = 2p(x^*)\Sigma - 2x^*
\]

where \(p(x^*) = p_1(x^*, x^*) = p_2(x^*, x^*)\). Recall \(x^*\) depends on \(W\) and \(L\), while \(W\), \(L\) and \(\Sigma\) depend on \(\alpha\) through expressions (4) and (5). The welfare measure is constructed by taking the probability that one of the two firms obtains a patentable innovation multiplied by the gross social value of the innovation minus the research costs. The optimal policy is partially characterized in the following two propositions.

**Proposition 1.** Collusive settlement, regular settlement, trial, and no entry all occur as the equilibrium litigation outcome of an optimal policy for certain environments.

**Proposition 2.** The litigation outcomes can be ranked in terms of the research investment that they induce. The ranking of the outcomes from lowest to highest investment is: collusive settlement, regular settlement, trial, and no entry.

**Proof.** Examples in the Appendix prove Proposition 1. To prove Proposition 2, take a fixed environment and suppose that condition (6) holds. From conditions (1), (2), and (3) it follows that the equilibrium litigation outcomes are ordered in terms of \(\alpha\). When \(\alpha < T/V_0\) the loser of the patent contest does not infringe. When \(T/V_0 \leq \alpha < [V_1 - V - 2T]/[V_1 - 2V_0]\) trial occurs. When \([V_1 - V - 2T]/[V_1 - 2V_0] \leq \alpha \leq [V_1 - V_0 - T]/[V_1 - V_0]\) regular settlement occurs. When \([V_1 - V_0 - T]/[V_1 - V_0] < \alpha\) collusive settlement occurs. If condition (6) does not hold then trial does not occur for any \(\alpha\). The ordering of the other three outcomes remains the same with the boundary between no entry and regular settlement occurring at \(\alpha = T/V_0\). This ranking in terms of \(\alpha\) can be related to the equilibrium research investment \(x^*\). From the payoff expressions \(W\) and \(L\) contained in Section IIA, one can see that \(W\) is decreasing in \(\alpha\) and \(L\) is increasing in \(\alpha\). From condition (9) it follows that \(x^*\) varies inversely with \(\alpha\). Therefore we have the ranking in the proposition. QED.

The optimization problem for the social planner is illustrated in Figure 11.1. A representative level curve for the social welfare function is shown with the constraint set. Expected social welfare rises with \(\Sigma\), and rises then falls in \(x^*\). The constraint set is determined by the equilibrium outcomes of the litigation settlement subgame. The figure shows the set is
not connected. The isolated points arise from the no entry and collusive settlement outcomes. Recall that payoffs are not sensitive to the probability of infringement for these outcomes because one of the players does not have a credible threat of going to trial. The lowest level of investment is achieved through collusive settlement because the patent is probably invalid, and payoffs for the winner and the loser of the patent contest are relatively close. The highest level of investment is achieved when the patent is probably valid and the loser of the contest does not enter. This is a winner-take-all scenario which induces heated research competition. The segment of the constraint set corresponding to equilibrium trial only exists when condition (6) is satisfied. This segment is downward sloping because a smaller value of $\alpha$ raises the value of a patent and equilibrium investment, but it also increases the probability of monopoly deadweight loss. Trial yields a higher investment than regular settlement simply because it arises from smaller values of $\alpha$ than settlement. The social optimum occurs at the tangency of one of the level curves with the constraint set (or else there is a corner solution).

The social planner faces a trade-off between a strong research incentive and a small deadweight loss. A patent that is certainly valid (or $\alpha < T/V_0$) induces the no entry outcome and tends to be optimal when the firms underinvest in research relative to the social optimum. Even though the deadweight loss is maximized, this policy may be optimal if the firms only
appropriate a small portion of the social value of the innovation. A patent that is certainly invalid (or $1 - \alpha < T/V_1 - V_0$) induces collusive settlement. The patentee does not have a credible threat of going to trial but can use a settlement license to raise industry profit. Despite the collusive nature of settlement this policy tends to be optimal when the social planner wants to discourage investment.\textsuperscript{10} Research investment may exceed the social optimum when the race to win the patent is the dominant factor affecting investment.

Regular settlement or trial offer intermediate levels of research incentive. Assume condition (6) is violated so that an equilibrium trial is not possible. The social planner may prefer regular settlement over collusion because it induces higher investment with the same deadweight loss. At the same time regular settlement induces lower investment than no entry and alleviates deadweight loss. If condition (6) holds, the social planner may prefer trial over the alternatives because it induces a level of investment that is intermediate between that induced by settlement on the one hand and no entry on the other. Getting this intermediate level of investment may be worth the sacrifice in trial costs. Furthermore, the gross social value of trial may be higher than no entry despite trial costs, because the possibility of invalidity may diminish the deadweight loss.

\section*{IV Antitrust and patent policy}
Ideally antitrust and patent policy should be coordinated to maximize expected social welfare. Antitrust policy towards settlement directly affects the output restrictions and deadweight loss attributable to licenses. It also affects the profitability of licensing and thus the incentives for research investment and the likelihood of settlement compared to trial. In this chapter I take antitrust policy as fixed and study the optimal choice of patent validity probabilities. Nevertheless it is possible to learn something about the interaction of the policies by determining the optimal response of patent validity probabilities to exogenous changes in antitrust policy.\textsuperscript{11}

The variables $S$ and $V$ capture the effect of antitrust policy. Social value, $S$, and industry profit, $V$, from licensing move in the opposite direction as antitrust policy changes. A stringent policy limits the ability of a licensor to introduce ancillary contract terms that restrict output. This corresponds to a relatively high value of $S$ and low value of $V$. A permissive policy has the opposite effect, leading to a low value of $S$ and high value of $V$.

A change in antitrust policy affects the settlement payoffs, but does not affect the trial or no entry payoffs.\textsuperscript{12} The absence of settlement licenses means that payoffs are unaffected given trial or no entry. When settlement occurs, the payoff to the winner of the patent contest grows and the social value of the innovation shrinks when antitrust policy is relaxed.
The reverse is true when antitrust policy is tightened. Because of the structure of the bargaining, the payoff to the loser of the patent contest is not affected by the antitrust policy. In addition to payoffs, a change in policy also shifts the boundary between trial and regular settlement. As the policy is relaxed trial becomes less likely. I will only consider marginal changes in antitrust policy and therefore ignore movement of this boundary.

The optimal probability of invalidity responds to antitrust policy only in the regular settlement case. Obviously, the optimal probability does not respond to a marginal in antitrust policy if the optimum induces trial or no entry in the litigation subgame. Antitrust policy has no effect. Furthermore, the optimal probability does not respond in the case of collusive settlement. The reason is simply that the optimal probability is indeterminate for collusive settlement. Any value \( \alpha \) such that \( 1 - \alpha > T/[V_1 - V_0] \) yields a collusive settlement. If the optimum induces regular settlement, then an exogenous change in antitrust policy provokes an offsetting change in patent policy. Specifically, if antitrust policy is relaxed, then the optimal probability of invalidity rises; if antitrust policy is tightened, then the optimal probability of invalidity falls. As a result, patent and antitrust policy must have offsetting effects on the profit of the winner of the patent contest.

Let me explain this result assuming a marginal tightening of antitrust policy. A tighter policy reduces deadweight loss from licensing. The social planner should respond by choosing higher research investment, because the marginal social value of investment has risen. The planner should notice that the tighter policy erodes the private incentive for investment, because industry licensing profits fall. Therefore, achieving higher investment requires a more secure patent and a lower probability of invalidity. I prove this result in the following proposition.

**Proposition 3.** When the optimal policy calls for regular settlement, the optimal probability of invalidity should be chosen to counteract the effect of changes in antitrust policy on the profitability of licenses.

**Proof.** Altering the antitrust policy moves \( V \) and \( S \) in the opposite direction. The first order condition for the social planner choosing \( x \) is given by \( Sp'(x) = 1 \). From this expression we find that \( dx/dS = -p'(x)/Sp''(x) \). Since the assumptions in the Appendix imply \( p'' < 0 < p' \), \( dx/dS > 0 \). Thus the optimal policy requires an increase in equilibrium investment when antitrust policy is made more stringent so that \( S \) rises. But the effect of a stricter policy is to decrease \( V \). This does not affect the profit of the loser of the patent contest from a settlement; it does reduce the settlement profit of the winner. Hence if \( \alpha \) is unchanged \( x \) would fall. The optimal policy
dictates that $\alpha$ must fall enough to more than offset the effect of $V$. Thus $\alpha$ must fall when the antitrust policy becomes more stringent. Similarly $\alpha$ must rise when the policy becomes more lax. QED.

V Nonobviousness standard

The critical hurdle in pursuit of a patent is Section 103 of US patent law – nonobviousness.\(^{14}\) The standard as interpreted in *Graham v. John Deere Co.*, 383 US 1 (1966), calls for three determinations: (1) the scope and content of the prior art, (2) the level of ordinary skill in the inventor’s art, and (3) the difference between the prior art and the claimed invention. If the difference is nonobvious to a person of ordinary skill in the art then Section 103 is satisfied.

Direct application of the test begets arcane technical issues that trouble judges and juries. To augment the test judges have fashioned various non-technical subtests.\(^{15}\) Roughly speaking the subtests fall into two categories. One category emphasizes the process of invention, e.g., failure by others, long-felt unmet nearly simultaneous independent invention (which works against the inventor), skepticism of experts, unexpected results, or movement of skilled researchers in a different direction. The other category emphasizes commercial aspects of the development of the invention, e.g., commercial success, immediate copying, or extensive licensing of the patent. There are no precise guidelines for integration of the subtests into the traditional three-part test. It is clear, however, that passing one or more of these subtests bolsters the inventor’s claim of nonobviousness. The effect of the subtests is to make the probability that the courts find an invention nonobvious depend on the attribute measured in the subtest. For example, the greater the commercial success of an invention the higher the probability of patent validity.

My goal here is to connect the optimal probability of validity in the model to these subtests of nonobviousness. I do that by examining the comparative statics of $\alpha^*$. I find that the optimal probability of validity is monotonic in certain parameters relating to the invention process, but not in parameters relating to commercial success. To obtain these results I restrict attention to the following research technology:\(^{16}\)

\[
p_i = \theta x_i (x_1 + x_2)^{\gamma - 1},
\]

where

\[
0 < \gamma < 1
\]

(11.11)

This functional form has some useful properties. Notice that the probability that one of the firms is successful is $\theta (x_1 + x_2)^\gamma$. Thus the variable $\gamma$ measures the elasticity of the probability of successful invention with
respect to total investment expenditure. A small value of $\gamma$ implies the probability of invention is not very sensitive to investment. Conversely, for large values of $\gamma$ the probability of success is very sensitive to investment. The scaling factor $\theta$ can be used to account for the intrinsic difficulty of achieving an innovation. It could also account for the patent approval decision by the Patent Office. An exogenous change in the approval rate would shift $\theta$.

The following proposition makes use of the functional form in (11) to provide a simple characterization of the optimal policy when it induces regular settlement.

**Proposition 4.** When the optimal policy induces regular settlement for the technology given in (11), the probability of invalidity, $\alpha^*$, solves:

$$2\gamma S = (1 + \gamma)V - \alpha V_0 + 2T \quad (11.12)$$

**Proof.** The relevant expressions for social and private payoffs are found in (4). Since $\Sigma^S$ is independent of $\alpha$, the first order condition from (10) gives $\gamma \theta (2x) \gamma^{-1} S = 1$. From (8) the private equilibrium value of $x$ satisfies $\theta(\gamma + 1)W^S + (\gamma - 1)L^S(2x)\gamma^{-1} = 2$. Combining these terms and substituting for $W^S$ and $L^S$ yields (12) which gives the unique maximum. QED.

The key result expressed in equation (12) is that the probability of invalidity $\alpha^*$ is inversely related to $\gamma$. In other words, the optimal policy is more favorable to the patentee when the elasticity of research technology is close to one. The reason is that equilibrium research investment tends to be too low relative to the social optimum for large values of $\gamma$ and too high for small values of $\gamma$. For example, suppose $\gamma$ is close to zero, then it is easy to see why there tends to be excessive investment. Since the probability that one of the firms will innovate, $\theta(2x^*)\gamma$, becomes close to the constant $\theta$, regardless of the total investment, the social planner wants to discourage research. In contrast, the firms want to win the patent race and thus overinvest. To reduce the firms’ investment incentive sufficiently the probability of invalidity must be high. When $\gamma$ is close to one, the equilibrium investment tends to be too low because the appropriability problem dominates, and the optimal probability of invalidity must be small.

This result supports the use of subtests based on attributes of the research process, but it does not clearly support any of the extant subtests. For example, the subtests relying on failure of others or long-felt unmet need could be associated with the model through $\theta$. The failure of others might be attributable to a low value of $\theta$. But from (12) we see that the optimal probability of invalidity is independent of $\theta$. At any rate the available subtests are not clearly linked to the elasticity of the research
technology. My result suggests a dichotomy between the serendipitous invention versus the Edisionian invention. When $\gamma$ is small it makes sense to say that a successful inventor was lucky. When $\gamma$ is large it makes sense to say that a successful inventor ‘perspired’; that the additional investment induced by a patent could have made the difference between success and failure.\textsuperscript{20}

Besides linking the nonobviousness standard to the elasticity of the research technology, expression (12) yields two other interesting comparative static results: first, $\alpha^*$ is decreasing in terms of the positive externality generated by the innovation, and second, the effect of the value of the innovation on $\alpha^*$ is ambiguous. It is evident from (12) that $\alpha^*$ is falling in $S$. For fixed licensing profit, $V$, a relatively high value of $S$ indicates that the innovation generates substantial social benefits, for example spillovers to other areas of research, that are not captured by an innovator. It would be straightforward and desirable to create subtest based on spill-over benefits created by an invention.

In contrast, the subtest based on the commercial success of an innovation is not supported by this analysis. Condition (12) can be used to study the effect of the value of the innovation on the optimal probability of validity. Here value refers to the commercial significance of the innovation, measured, for example, by the intercept of linear demand curve. An upward shift in demand raises $V_0$, $V$, and $S$. Examples can be adduced in which $\alpha^*$ either rises or falls. The ambiguous effect of the value of the innovation on the probability of invalidity arises because both the socially optimal investment, and the private equilibrium investment rise with the value of the innovation. It may be the case that the equilibrium investment rises too fast relative to the social optimum, and an increase in the probability of invalidity would be required. Or conversely, the equilibrium investment may rise too slowly which requires a decrease in the probability of invalidity.\textsuperscript{21}

VI Conclusion

In this chapter I undertake a normative analysis of new product innovation in a model in which the government chooses the optimal probability of patent validity. The major goals of the chapter are to explicate the role of antitrust policy and costly litigation in the patent system, and to determine the optimal implementation of the nonobviousness standard of patentability. I find that, in a model with complete information, equilibrium patent trials occur for certain probabilities of validity because of asymmetric stakes in the litigation created by the patent licensing process. But trials vanish from equilibrium in this model if the antitrust policy governing licensing is too lenient. An optimal policy might result in trial, because
the intermediate probabilities of patent validity which lead to intermediate levels of research effort also give rise to trial. It may be the case that weak patents which lead to settlement result in too little research, and strong patents which lead to unchallenged monopoly result in too much research; consequently, equilibrium trials are tolerated under the optimal policy.

The nonobviousness standard is intended to limit patent grants to inventions that represent a significant advance over previous technology. To implement this standard the courts have relied on various tests related to the research process as well as tests related to commercial success of the inventions. I assume a particular constant elasticity research process and show that the optimal probability of validity increases in that elasticity.

In contrast, the linkage between commercial success and nonobviousness may be undesirable, since the optimal probability of validity is not monotonic in the commercial value of the innovation.

Notes

1. Professor of Law, Michaels Faculty Scholar, Boston University School of Law. Thanks to Ed Prescott, Herb Mohring, Bob Marshall, and Paul Joskow for helpful comments. Previous versions of this chapter were distributed as: ‘Designing an Optimal System of Patent Litigation’ (July 1987), ‘Optimal Patent Litigation’ (December 1992), and ‘Patent Litigation, Licensing, and the Nonobviousness Standard’ (July 1995).


3. A survey of patent licensors revealed that 27 per cent of licenses covered products for which there is no close substitute. See Caves, Crookell and Killing (1983).

4. I am ignoring the possibility that the patentee licenses competing manufacturers to mitigate the impact of the hold-up problem on purchasers. This theory is explored by Shepard (1987) and Farell and Gallini (1988).

5. There are settlement cases concerning sham process patents, and a case involving the novelty of a drug, in which antitrust charges have been sustained against the licensor and licensees. Hence it may be the case that the factors that determine the probability of validity also determine the joint settlement profit. I disregard this possibility and assume that V is constant. See Priest (1977), for a discussion of patent licenses used to facilitate cartels.

6. For a more general treatment of the settlement of patent litigation, see Meurer (1989).

7. The loser of the patent contest does not have a credible threat unless there is some barrier to entry to production and sales besides the patent. There would be no profits and no incentive to go to trial otherwise. If more than two firms can compete after the patent is invalidated then V_0 could be reinterpreted as per firm profit for a market with n firms. Obviously, V_0 would be smaller and credibility would be a bigger problem for the loser of the patent contest.

I also assume there are only two active firms at the litigation and research stages. Multiple potential infringers create modeling problems associated with the public good nature of patent litigation and multilateral bargaining. Assuming one potential infringer avoids these complications.

8. A payment by the patentee in exchange for a promise by its rival to withdraw from the market and never challenge the patent’s validity would eliminate equilibrium trials. So would a promise by the patentee to exit the market or a merger between the firms.
More subtle means of achieving the same effect are possible through restraints in license agreements. If these tactics are precluded by antitrust law, then equilibrium trials are a possibility.


10. The use of weak or insignificant patents to cartelize an industry through licensing has been successfully attacked under the antitrust laws. See Priest (1977). In the setting of this model it would be difficult for the government to win a case because it would require a showing that the license was not in settlement of a bona fide dispute. If the antitrust authorities could distinguish and prohibit collusive settlement licenses, then the payoffs would change so that \( W^C = L^C = V_0 \) and \( \Sigma^C = S_0 \). The analysis of the optimal policy would not change significantly.

11. For other work studying the patent and antitrust interface, see Kaplow (1984), Chang (1995), and Green and Scotchmer (1995).

12. Antitrust policy can affect the litigation and no entry payoffs. The payoff to the monopolist, \( V_1 \), can be affected especially by Section 2 of the Sherman Act, whereby limits are placed on the ability of monopolists to exploit their market power. I only consider antitrust policy toward settlement licenses.

13. If the licensing surplus were shared equally the result in the proposition is not significantly affected. An increase in \( V \) would raise the payoffs to the winner and loser equally. For typical research technologies this will lead to greater research investment ex ante, and the same proof applies.

14. Non-obviousness ‘appears to be the predominant ground in court decisions of invalidity’. Kitti (1979) at p. 56.


16. The parameter \( \theta \) must be chosen to ensure that the probabilities \( p_1 \) and \( p_2 \) are well defined.

17. From (8) the private equilibrium investment is \( (1/2)(\theta/2)(W + \gamma (W + L) − L)^{1/(1-\gamma)} \). From (10) the socially optimal investment is \( (1/2)(\gamma S)^{1/(1-\gamma)} \). So private investment is too high if and only if \( \gamma < [W − L]/[2S − W − L] \).

18. The equilibrium probability that one of the firms will get a patent is given by \( \theta ([\theta/2](W + \gamma (W + L) − L)^{1/(1-\gamma)} \). This has a limit of \( \theta \) as \( \gamma \) goes to zero.

19. The variable \( \alpha \) is determined in part by the rate of patent approval in the Patent Office. The fact that \( \alpha^* \) is independent of \( \theta \) means that exogenous changes in the behavior of the Patent Office should not influence the probability of validity. Assuming a large random element to patent approval by the Patent Office is reasonable because of the limited information and resources available for processing applications, and because of the ex parte nature of the proceedings. Nevertheless, it would be valuable in future work to allow both the Patent Office and the courts to control the probability of validity.

20. Implementing a subtest based on the elasticity of the research technology would be complicated. One approach would rely on statistical measures of the elasticity derived from other research projects in the same area of technology.

21. This conclusion differs from Kitch (1977) who advocates a weak standard of nonobviousness and believes that commercial success should be used to infer nonobviousness. His claims cannot be evaluated in the context of this model since he addresses a research process of cumulative innovation that has no counterpart in this model. Merges (1988) dissents from Kitch and reaches conclusions complementary to mine.

References


Cases


Statutes


Appendix

The model features a typical patent contest in which \( p_i(x_1, x_2) \) is the probability that firm \( i \) receives a patent on a new product invention. I assume that (1) the probabilities \( p_1 \) and \( p_2 \) are twice continuously differentiable, non-negative and their sum is less than or equal to one, (2) the technology available to firms 1 and 2 is identical, thus, \( p_1(x_1, x_2) = p_2(x_2, x_1) \) for any values of \( x_1 \) and \( x_2 \), (3) the probability of success is strictly increasing in own investment and strictly decreasing in the other firm’s investment, and (4) \( \partial^2 p_i / \partial x_1 \partial x_2 < 0 \) at \( x_1 = x_2 \). Assumption (4) rules out large spillovers and requires that the marginal gain in the probability of a patent from own investment is declining as the other firm’s investment increases. One last assumption (5) is used to assure the second order conditions are satisfied: for \( W \geq L \geq 0 \), I assume \( W \partial^2 p_1 / \partial x_1 \partial x_1 + L \partial^2 p_1 / \partial x_2 \partial x_2 < 0 \).

I first show that a unique equilibrium pair \((x_1, x_2)\) exists for the patent contest with payoffs given in expression (8) in the main text. From Theorem 7.1 in Friedman (1977) I have existence. I limit the choice of \( x_1 \), and \( x_2 \) to some interval making the strategy set compact. This can be done without loss of generality since a choice of \( x_1 > V_1 \) would be unprofitable given any \( x_2 \) and could not be part of a Nash equilibrium. Furthermore, the payoff functions are strictly quasiconcave by (5), so existence is assured.

The equilibrium in research expenditures is unique if the best reply mappings are contractions according to Theorem 7.7 in Friedman (1977). Showing that the absolute values of the slopes of the reaction functions are less than one is sufficient to show uniqueness. The reaction functions are given implicitly by:

\[
Wp_1^1 + Lp_2^2 = 1 \\
and \\
Lp_1^2 + Wp_2^1 = 1
\]

(11.A1)

Applying the implicit function theorem to the first order conditions gives a slope of:

\[
- \frac{Wp_1^1 + Lp_2^2}{Wp_1^{11} + Lp_2^{22}}
\]

(11.A2)

Thus, the reaction function is a contraction if:

\[
|Wp_1^1 + Lp_2^2| > |(W + L)p_1^2|
\]

(11.A3)

This inequality does not follow from the assumptions on \( p \) and is assumed to hold guaranteeing uniqueness.
Using the implicit function theorem and (A1) one can write \( x_1 = f_1(W,L) \) and \( x_2 = f_2(W,L) \). (A3) assures the implicit function theorem can be used. By symmetry \( x_1 = x_2 = f(W,L) \). Expression (9) in the main text holds through application of the implicit function theorem to (A1) through the use of assumptions (4) and (5).

Finally, I use the research technology in expression (11) in the text to provide four examples in which each of the litigation outcomes is part of the socially optimal patent policy. Suppose \( \gamma = 1/2, V_0 = 1, V = 3, V_1 = 4, S_0 = 10, S = 8, S_1 = 6, \) and \( T = 0 \). Then \( \alpha^* = 0.384 \) and trial occurs in equilibrium. Expected social welfare under the optimal policy is approximately 2.54 per cent higher than in the best settlement outcome, approximately 24.62 per cent higher than in the case of no entry and approximately 32.92 per cent higher than in the case of collusive settlement. The choice of \( T = 0 \) is not required; there is sufficient continuity in the problem that trial still occurs in the optimal policy for sufficiently small \( T \).

In the next example the only change from above is \( S = 9.5 \) and \( S_1 = 9 \).

Then \( \alpha^* = 0 \) and no entry is socially optimal. In this case for \( T = 0 \) trial is the same as no entry, but for any \( T > 0 \), trial is strictly worse. Expected social welfare under the optimal policy is approximately 32.72 per cent higher than in the best regular settlement outcome, and approximately 74.55 per cent higher than in the case of collusive settlement.

For the third example let \( \gamma = 1/2, T = 0, V_0 = 1, V = V_1 = 3, S_0 = 5, \) and \( S = S_1 = 4 \). Then \( \alpha^* = 0.25 \) and regular settlement is socially optimal. Trial is no longer a feasible outcome. Expected social welfare under the optimal policy is approximately 1.59 per cent higher than in the case of no entry, and approximately 16.36 per cent higher than in the case of collusive settlement.

For the final example use the values above except \( V = V_1 = 5, S_0 = 6, \) and \( S = S_1 = 5 \). Then \( \alpha^* = 1 \) and collusive settlement is socially optimal. Here trial is not feasible, and the best regular settlement is equivalent to collusive settlement. Expected social welfare under the optimal policy is 32 per cent higher than in the case of no entry.