Shared Rights and Technological Progress*

Matthew Mitchell[†] and Yuzhe Zhang[‡]

October 24, 2012

Abstract

We study how best to reward innovators whose work builds on earlier innovations. Incentives to innovate are obtained by offering innovators the opportunity to profit from their innovations. Since innovations compete, awarding rights to one innovator reduces the value of the rights to prior innovators. We show that the optimal allocation involves shared rights, where more than one innovator is promised a share of profits from a given innovation. We interpret such allocations in three ways: as patents that infringe on prior art, as licensing through an optimally designed ever-growing patent pool, and as randomization through litigation. We contrast the rate of technological progress under the optimal allocation with the outcome if sharing is prohibitively costly, and therefore must be avoided. Avoiding sharing initially slows progress, and leads to a more variable rate of technological progress.

JEL Classification Numbers: D43, D82, L53, O31, O34

Keywords and Phrases: Cumulative Innovation, Patent, Licensing, Patent Pool, Litigation.

*We thank B. Ravikumar, Hugo Hopenhayn and participants at seminars at the Einaudi Institute for Economics and Finance and Queens University for helpful comments.

[†]Rotman School of Management, University of Toronto

[‡]Department of Economics, Texas A&M University

1 Introduction

An important question in the economics of innovation is how to structure rewards for innovators. A long list of authors have argued that because an innovation's quality is unobservable, rewards must take the form of rights to profit from the innovation, rather than a simple procurement contract.¹ This manifests itself in the public policy through patents, and in private contracts through licensing agreements that pay royalties.

A more recent debate has addressed how to reward cumulative innovations. When one innovation will be built upon by future innovators, how should the rights of the earlier innovator be balanced against the rewards of those who come later? In this paper we address the role of sharing across innovators in the efficient reward of innovators under incomplete information of the sort that motivates patents and royalty payments. We show that, in a contracting environment that supports arbitrary ex ante sharing agreements, the optimal allocation involves shared rights: history does not imply a unique firm with rights to the profit flows that arise from the current state of the art. We then contrast that allocation with regimes where institutional arrangements do not allow for sharing, and show that lack of sharing leads to more variable and possibly slower technological progress.

Sharing in our model relates to commonly observed practices in patents and licensing. Patents offer protection in two ways: (1) competing innovations might be excluded by being deemed unpatentable; and (2) a later innovation might be patentable but still infringe on the earlier patent. In the second case, rents from the new innovation may be shared through a licensing contract where both firms gain a fraction of additional profit generated by their joint product. In cases where many innovators contribute to a common technology, more contributors are eligible to share the profits. For instance, in a patent pool, many firms can potentially get a share of the future profits from the patents under license. The recent literature on weak patent rights argues that randomization through litigation is a pervasive feature of current patent policy. Randomization can be

¹John Stuart Mill (1883) wrote that patents are useful “because the reward conferred by it depends upon the invention's being found useful, and the greater the usefulness, the greater the reward.”

viewed as shared rights, in the sense that several patent holders have a probabilistic claim to future profits. Our model, then, provides a rationale for these observed practices.

Our model features a sequence of innovators who make quality improvements on one another. An innovator endowed with an idea can develop a product with a quality that is one unit greater than the previous maximum quality after paying a one-time cost to develop his idea. This cost is innovator-specific but is drawn from a common distribution. To maximize welfare, a social planner would like to implement any idea whose cost is lower than its social benefit. The central question, then, is how to reward innovators so that they are willing to pay the costs.

Without any sort of frictions, rewarding innovations is a trivial problem: each innovator can be compensated for his contribution through a monetary payment. However, neither an innovator's expenditure toward the improvement nor the quality of his product can be verified by the planner in our model. This moral hazard problem leads to a situation where an innovator cannot be compensated through a monetary payment. Instead, the planner rewards the innovator with the opportunity to profit from innovations that embody his contribution. In contrast to a monetary payment, this reward ensures that the innovator has an incentive to pay the cost to develop his idea, because the qualities and hence profits of future products are contingent on the innovator's contribution.

Since we study a model of cumulative innovation, there is scarcity in this form of reward: market profits are limited and must be divided across innovators to provide incentives. The planner must decide how to divide an innovator's reward between marketing his own innovation, perhaps excluding future innovators in the process, and allowing the innovator to share in profits from future innovations that build on his original contribution. If the planner excludes future innovators, this increases the incentive to the current innovator by lengthening the time during which the innovator can market his innovation. In Hopenhayn et al. (2006), sufficient conditions are developed such that these exclusion rights are the only reward the innovator receives; when an improvement arrives, the existing innovator's right to profit ends. By contrast, the planner in our environment always gives innovators a stake in future innovations; this also means that multiple firms have rights to the profit flows at a moment in time.

One can equally interpret the planner’s allocation in our environment as the outcome of a license that maximizes ex ante surplus among innovators, or as arising from a normative policy design problem. As such, one can interpret the shared rights as reflecting licensing features such as patent pools, or policy choices such as weak patent rights. When shared rights take the form of a patent pool, the proceeds from the state of the art are divided among the innovators according to a preset rule: a constant share is given to every new innovator when he is allowed into the pool, while old innovators’ shares are reduced proportionally to make room.²

The optimal ex ante licensing contract brings to light a new sense in which a patent pool might be “fair, reasonable and non-discriminatory” (FRAND). FRAND licensing is a standard metric for judging patent pools, but existing interpretations of FRAND pertain to the patent pool’s treatment of the users of the pooled patents, which in our model are consumers. Our model is silent on this issue; here the focus is on the rules by which innovators are allowed to join the pool. There is a preset price for an innovation to join, which is common to all potential innovators, and therefore the formation of the pool can be interpreted as FRAND with respect to new arrivals. One can interpret a policymaker’s role here as to ensure treatment is FRAND even when contracts cannot be completely written ex ante.

The second interpretation of shared rights in our model is through a lottery among the risk-neutral innovators who share the rights, where a winner keeps all future profits promised to both winner and loser. This matches the notion of weak patent rights in the literature, where the randomization is commonly interpreted as litigation.

The lottery interpretation sheds light on both the cost and the benefit of shared rights. On the cost side, shared rights may require specific institutions for conflict resolution, since sharing may lead to contracting issues or legal conflict. Because the innovators are not neutral with respect to the outcome of the lottery, there is an incentive to spend resources to rig the lottery or lobby the planner to report that the lottery came out in their favor. Interpreting the lottery as litigation, the court allows the innovators

²As well, sharing may appear as a “grant-back” licensing contract, where the licensee grants rights on his future innovation to the licensor.

to hire lawyers and experts to increase their chances of winning the lawsuit. Such spending can be interpreted as a wasteful cost of sharing. On the benefit side, sharing is a convexification device for the planner. Without it, the planner must choose to grant rights entirely to one innovator or another, rather than choosing something in between. Viewed in this light, the policy with sharing naturally leads to smoother levels of total rights granted than the policy without sharing. With sharing, improvements arrive at a constant rate and different arrivals generate equal expected net social benefits.

Our sharing contract embodies a great deal of ex ante licensing agreements that may be impossible to negotiate in many environments. Rather than modeling the imperfections of contracting explicitly, our model allows us to assess the impact if the planner must avoid sharing completely. Without sharing, rights never come into conflict, in the sense that when the next innovation is developed, there is unambiguously no property right left for past innovators. In such cases, the allocation of rights is distorted: the planner bypasses some higher net benefit ideas at the beginning and implements other less attractive ideas later on; technological progress could be in perpetual cycles, in contrast to the smooth progress that results from allocations with sharing.

Related literature Our paper relates to several strands of the patent literature. It takes up, in the spirit of Green and Scotchmer (1995), Scotchmer (1996), O’Donoghue et al. (1998), O’Donoghue (1998), and Hopenhayn et al. (2006), among others, the allocation of patent rights when early innovations are an input into the production of subsequent innovations, and therefore rights granted to a later innovator reduce the value of rights granted to an earlier innovator. We study, in a model where patents are the result of information frictions as in Scotchmer (1999), the sharing question of Green and Scotchmer (1995) and Bessen (2004). We study both the evolution of sharing and the implications of the lack of sharing. Our optimal policy generates a patentability requirement (a minimum requirement for a patent to be awarded) in the spirit of Scotchmer (1996) and O’Donoghue (1998). The policy also implies situations where a new innovation is awarded protection but must share with prior art, which can be naturally interpreted as patent infringement, a feature of policy emphasized

in O'Donoghue et al. (1998). Our model weds the sequential innovation structure of Hopenhayn et al. (2006) with an underlying idea process in keeping with Scotchmer (1999). We show how the combination leads to optimal allocations featuring sharing, patentability, and infringement from a problem with explicit incomplete information, and therefore provide an informational foundation for the optimal policies studied in many papers on cumulative innovation.

Our model also provides a new interpretation of patent pools. Existing models, such as Lerner and Tirole (2004) and Gallini (2012) treat patent pools as similar to mergers, focusing on the conduct of the pool vis-a-vis its customers. Lerner et al. (2007) extend this approach to analyze the sharing rules that underlie joint production in a pool. Our interpretation offers a different view: patent pools split rights among competing innovators in a way that gives incentives for new members to produce innovations that enhance the pool. In that sense, our work connects the sharing rules for patent pools with the incentive-to-innovate issue at the heart of the optimal patent literature.

Many environments take the probabilistic view of patent protection. This notion of weak patent rights is introduced in Shapiro (2003), and the idea of patents being probabilistic is reviewed in Lemley and Shapiro (2005). Whereas game theoretic models of litigation often take the outcome of litigation to be random,³ our paper takes a different view: rather than assume patent rights are weak, we provide foundations for such environments, addressing the question of why a planner would choose an arrangement where patent rights are probabilistic.

Our model is fundamentally driven by a desire to solve moral hazard problems between a group of agents. Classic models of moral hazard in teams date back at least to Holmstrom (1982). Bhattacharyya and Lafontaine (1995) address sharing rules with double-sided moral hazard in the context of joint production between two agents, for instance between a franchisor and franchisee. Our model extends these ideas to a sequence

³Examples are numerous. First, papers that are more general than models of patent litigation include Bebchuk (1984), Nalebuff (1987), and Posner (2003). These papers study random litigation mostly in the context of incentives to settle pretrial. Second, papers on patent litigation include Meurer (1989), Choi (1998), and Aoki and Hu (1999). Chou and Haller (2007) follow in the spirit of the weak patent literature and consider a sequential environment, as we do.

of innovators and adds a new trade-off that comes out of the dynamics: the planner can offer rewards either through sharing or through exclusion of future innovators. We therefore are able to assess the benefits of rules that involve sharing.

Organization The remainder of the paper is organized as follows. The environment is described in Section 2. Section 3 describes optimal policies if sharing is costless. Section 4 discusses in more detail the interpretation of such policies as generating conflict. Section 5, motivated by these interpretations, takes the alternate view, constructing optimal policies without sharing. In Section 6 we compare and contrast the outcomes under the two scenarios and consider the possibility that the first innovator needs a greater reward, for instance because of the high cost of the initial innovation, or because it is less profitable than the follow-ups. Section 7 concludes, and we provide the proofs of all the results in an appendix.

2 Model

2.1 Environment

The environment has an infinite horizon of continuous time. There are many *innovators* (who we sometimes call firms) and a patent authority (who we sometimes call the planner). Everyone is risk-neutral and discounts the future with a constant discount rate r . Firms maximize profits, while the planner maximizes the sum of consumer surplus and profits. We call an opportunity to generate an innovation an *idea*. Ideas arrive with Poisson arrival rate λ . The idea comes to one of a continuum of innovators with equal probability, so that with probability one each innovator has at most one idea, and therefore an arrival can be treated as a unique event in the innovator's life. The arrival of the idea is private information of the innovator.

When the idea arrives, the innovator draws a cost of development c from a continuous distribution $F(c)$ with density $f(c)$. The draw of c is also private information of the innovator. To ensure that the planner's problem is convex, we impose the usual reverse-hazard-rate assumption on f .

Assumption 1 $\frac{f(c)}{F(c)}$ weakly decreases in c .

Investment of c leads to an *innovation*. Innovations generate higher and higher quality versions of the same good. Ideas are perishable, so the investment must be made immediately after the idea arrives, or it is lost. Every innovation generates a product whose quality is one unit greater than the previous maximum quality. In other words, the n th innovation is a product of quality n . Our notion of an idea matches that in Scotchmer (1999), which is a combination of a value and a cost required to achieve the value; here the value of an idea is always one. In Appendix B we show that our results extend to the case where value varies with cost; we also discuss how sharing can arise when innovation size is endogenously chosen by the innovator. We focus on the case of a fixed innovation size for analytic tractability in the body of the paper.

There is a single consumer who demands either zero or one physical unit of the good. The benefit to the consumer of the good when quality is n and price is p is $n - p$. The consumer has an outside option of zero. Firms have no cost of production. This implies that the exclusive rights to sell one unit allow the rights holder to make n units of profits. As in Hopenhayn et al. (2006) we assume that there are no static distortions. Since we are interested in the conflict that arises between early and late innovators when rights are scarce, we focus on the role of dynamic forces exclusively.

So that the optimal reward structure is not simply a prize system, we assume there is moral hazard. Both the investment and the quality of the good are assumed to be unobserved by the planner. As in the literature on optimal patent design with asymmetric information, the moral hazard makes pure transfer policies impossible, since innovators would always prefer to underinvest and collect the prize.⁴ We therefore focus on policies that reward innovators with the opportunity to market their innovations, which mimic the observed practice of patent policies.

⁴Without asymmetric information of any type, the optimal policy would be simple: transfer c to the inventor if c is less than the social gain from the innovation. If c were unobserved but moral hazard were absent, the planner would then choose a cutoff $\bar{c}$ and offer a transfer of $\bar{c}$ to all innovators. The only difference would be that inframarginal cost types would earn an information rent. In either case, the problem becomes one of public finance: how to raise the resources.

In our model, a firm with the exclusive rights to sell the leading-edge product knows the quality of the product because it is equal to his profit. The planner may ask the firm to report the quality, to cross-check past innovators' contributions. However, we assume that, although the quality and market profits are observable by the firm, they are not verifiable. We further assume, as in Scotchmer (1999) and much of the patent literature, that the consumer is a passive player who cannot be made to give a report of quality.⁵ These informational limitations imply that a firm will not face punishment by the planner (or legal action by other firms) when he claims an innovation but actually does not develop it. We abstract from cross-reporting followed by punishment or legal action because cross-reporting contracts are fraught with bribery concerns (see Hopenhayn et al. (2006)) and are studied in more detail in a literature following Kremer (1998).

2.2 Planner's Choices

At every point in time, the planner determines the identity of an exclusive rights holder for the leading-edge product. The planner can charge a fee in exchange for the rights granted; fees collected are rebated to the consumer lump-sum. If an innovator claims to receive an idea at date t , the planner asks him to report his cost c and makes a recommendation to him with a cutoff $\bar{c}_t$: pay the investment cost c and innovate if $c \leq \bar{c}_t$; do not invest otherwise. If the innovator claims $c \leq \bar{c}_t$, he is eligible to receive exclusive rights (and pay the fees) as specified by the history-dependent policy. If he does not, there is no further reward or punishment.⁶ Appendix B shows that the recommendation for investment must follow a cutoff rule in any incentive compatible contract.

More specifically, let n_t be the total number of innovators who claimed innovation and let ϕ_t be the fee. An indicator function $I_t : \{1, 2, \dots, n_t\} \rightarrow \{0, 1\}$ determines the exclusive rights holder, i.e., innovator i holds the exclusive rights if and only if $I_t(i) = 1$.

⁵Kremer (2000), Chari et al. (2011), Weyl and Tirole (2011) and Henry (2010) study situations where the planner uses market signals.

⁶There can be neither reward nor punishment for an innovator if he does not invest. If there is reward, an innovator with no idea would pretend to have one. If there is punishment, an innovator with a high-cost idea would hide the arrival of his idea.

At most one of $\{I_t(i) : i = 1, 2, \dots, n_t\}$ can be 1 and the planner may allow no innovator to profit in a given instant by choosing $I_t(i) = 0$ for all i . We allow the planner to condition $\bar{c}_t$, I_t , and ϕ_t on the history at time t . Because we ruled out reports of quality or market profits in the last section, history only records the number of arrivals, the arrival times, and reported types of prior innovators. We show in Appendix B that conditioning on past reported costs (beyond whether or not c is below $\bar{c}_t$) will not improve welfare. Because such type dependencies do not improve welfare but do complicate notation, we focus in the body of the paper on a simpler contract that only conditions on whether or not c is below $\bar{c}_t$.

We study how the planner's policies affect innovators' incentives next.

2.3 Incentive Compatibility

Suppose innovator i receives an idea at date t . He has three options: 1) claim a cost below $\bar{c}_t$ and invest, 2) claim a cost below $\bar{c}_t$ but shirk, and 3) not claim a cost below $\bar{c}_t$ and receive nothing. If innovator i chooses option 1 and is assigned rights at date $s \geq t$ (i.e., $I_s(i) = 1$), his profit flow is n_s minus the fees ϕ_s ; otherwise, it is zero. Therefore the present discounted payoff of option 1 is

$$E \int_t^\infty e^{-r(s-t)} I_s(i) (n_s - \phi_s) ds - c,$$

where the expectation is taken with respect to uncertain future states n_s . On the other hand, if the agent shirks and chooses not to pay the cost of developing the innovation after claiming a cost below $\bar{c}_t$, the future states will all be lower by one, and therefore the payoff to shirking is

$$E \int_t^\infty e^{-r(s-t)} I_s(i) (n_s - 1 - \phi_s) ds.$$

That difference between options 1 and 2 is simply $d_1 - c$, where

$$d_1 \equiv E \int_t^\infty e^{-r(s-t)} I_s(i) ds,$$

is the expected discounted *duration* of rights. Note that d_1 can never exceed r^{-1} ; were the planner to offer full monopoly to the innovator forever, he would receive $d_1 =$

$\int_t^\infty e^{-r(s-t)} ds = r^{-1}$ and any cost type less than r^{-1} would invest. If the planner has made prior promises, d_1 will be constrained by those promises and be less than r^{-1} , so that not all cost types smaller than r^{-1} will invest. This is the fundamental trade-off in the model: making promises implements more ideas but restricts the feasible set of future payoffs.

Definition 1 *The planner's policy is **incentive compatible** if an innovator with $c \leq \bar{c}_t$ prefers to invest (option 1), and an innovator with $c > \bar{c}_t$ prefers no rights and no fees (option 3).*

Lemma 1 *The contract is incentive compatible if and only if*

$$(i) \quad \bar{c}_t \leq d_1;$$

$$(ii) \quad E \int_t^\infty e^{-r(s-t)} I_s(i)(n_s - \phi_s) ds - \bar{c}_t = 0.$$

The proof of Lemma 1 is briefly explained as follows. Part (i) states that an innovator with cost $\bar{c}_t$ will invest. Part (ii) states that he is indifferent between option 1 and option 3. Both (i) and (ii) are necessary for incentive compatibility. To see they are also sufficient, we substitute (ii) into (i) and obtain

$$E \int_t^\infty e^{-r(s-t)} I_s(i)(n_s - 1 - \phi_s) ds \leq 0,$$

which states that the innovator's payoff is nonpositive if he chooses option 2.

2.4 Planner's Objective

Although the firm's private benefit from the innovation is d_1 , the social value of an innovation is r^{-1} , since the quality improvement of one unit is always generating profits for some firm (if exclusive rights are granted) or consumer surplus (if no innovator is granted exclusive rights and price is zero). Therefore the planner's expected payoff from an idea with cutoff $\bar{c}$ is

$$R(\bar{c}) \equiv \int_0^{\bar{c}} (r^{-1} - c) f(c) dc.$$

Since no innovator ever develops an idea with $c > r^{-1}$ (even if offered the maximum feasible $d_1 = r^{-1}$), we can assume that the support of the density $f(\cdot)$ is $[0, r^{-1}]$ and $R(\cdot)$ is increasing without loss of generality.

Note that the fees do not enter the social welfare function: they increase the consumer surplus but decrease firms' profits. Hence part (ii) in Lemma 1 is not a real constraint: the planner can always adjust ϕ_s to satisfy (ii) without affecting social welfare. On the other hand, since $R(\bar{c})$ is increasing in $\bar{c}$, we may assume $\bar{c} = d_1$ without loss of generality. To summarize, the planner's problem is

$$\begin{aligned} \max_{\{I_t; t \geq 0\}} \quad & E \left[\sum_{i=1}^{\infty} e^{-rt_i} R(\bar{c}_{t_i}) \right] \\ \text{subject to} \quad & I_t = 1 \text{ for at most one innovator, } \forall t \geq 0, \\ & \bar{c}_{t_i} = E \left[\int_{t_i}^{\infty} e^{-r(s-t_i)} I_s(i) ds \right], i = 1, 2, \dots, \end{aligned}$$

where t_i is the arrival time of the i th idea, and $\bar{c}_{t_i} = E \left[\int_{t_i}^{\infty} e^{-r(s-t_i)} I_s(i) ds \right]$ is both the cutoff cost and the promised duration if innovator i claims a cost below $\bar{c}_{t_i}$.

2.5 A Recursive Formulation

At date t , the future duration promised to innovator i , who arrives before t , is

$$E \left[\int_t^{\infty} e^{-r(s-t)} I_s(i) ds \right].$$

The past history can be succinctly summarized by the sum of durations promised to all prior innovators (i.e., $\sum_{i=1}^{n_t} E \left[\int_t^{\infty} e^{-r(s-t)} I_s(i) ds \right] \equiv d$); the duration available to subsequent innovators is $r^{-1} - d$. History matters only because this is a full-commitment problem, and hence the planner must deliver on this promise. A greater promise limits what can be offered to future innovators, and this is the fundamental scarcity that leads to conflict and makes the solution differ from the first best. Summarizing histories in this way allows us to characterize optimal policies recursively in the next section.

3 Optimal Policies

Suppose the planner finds herself with an outstanding duration promise of d at time t . The planner can offer rights $y \in [0, 1]$ to the existing rights holders in the intervening instants until the next arrival. The planner may freely dispose of instants by choosing $y < 1$, leaving $1 - y$ unassigned to any one, and letting the leading-edge product be sold at the marginal cost of zero. If a newly arrived innovator claims rights and pays the fees, then he is offered a duration promise of d_1 . In this event, the incumbents' promise is changed to d_0 . Both d_1 and d_0 are contingent on d . The planner can also adjust the incumbents' promise when no idea arrives (i.e., $\dot{d} \equiv d'(t) \neq 0$) or when an innovator arrives but does not invest (i.e., change the incumbents' promise to $\hat{d} \neq d$). However, Lemma 3 in Appendix B shows that neither instrument is useful, which allows us to impose $\dot{d} = 0$ and $\hat{d} = d$ below.

Recall from the previous section that a new innovator claims rights if and only if his cost is below d_1 . Such an innovator arrives with Poisson rate $\lambda F(d_1)$ and let us denote his arrival time as τ . The following promise-keeping (PK) constraint ensures that the planner delivers the promised d to the incumbents:

$$d = E \left[\int_t^\tau e^{-r(s-t)} ds \right] y + E \left[e^{-r(\tau-t)} \right] d_0,$$

where $E \left[\int_t^\tau e^{-r(s-t)} ds \right] y$ is the expected rights offered to the incumbents from t to τ , and d_0 is the rights offered from τ onwards. The discount factor $E \left[e^{-r(\tau-t)} \right]$ brings d_0 , which is discounted to τ , back to time t . After some algebra, the PK constraint becomes⁷

$$d = \frac{1}{r + \lambda F(d_1)} y + \frac{\lambda F(d_1)}{r + \lambda F(d_1)} d_0. \quad (1)$$

⁷Because the density function for τ is $\lambda F(d_1) e^{-\lambda F(d_1)(x-t)}$ at $x \geq t$,

$$\begin{aligned} E \left[\int_t^\tau e^{-r(s-t)} ds \right] &= \int_t^\infty \left[\int_t^x e^{-r(s-t)} ds \right] \lambda F(d_1) e^{-\lambda F(d_1)(x-t)} dx = \frac{1}{r + \lambda F(d_1)}, \\ E \left[e^{-r(\tau-t)} \right] &= \int_t^\infty e^{-r(x-t)} \lambda F(d_1) e^{-\lambda F(d_1)(x-t)} dx = \frac{\lambda F(d_1)}{r + \lambda F(d_1)}. \end{aligned}$$

Since $y \leq 1$, the PK constraint (1) is equivalent to an inequality⁸

$$d \leq \frac{1}{r + \lambda F(d_1)} + \frac{\lambda F(d_1)}{r + \lambda F(d_1)} d_0. \quad (2)$$

If (2) is slack, then $y = (r + \lambda F(d_1))d - \lambda F(d_1)d_0 < 1$: some intervening instants are not assigned to any firm in order to satisfy the PK constraint (1).

The planner's value, $V(d)$, is the maximized discounted sum of social surpluses of all innovations in the future. It satisfies

$$\begin{aligned} V(d) &= E[e^{-r(\tau-t)}] \left(\int_0^{d_1} (r^{-1} - c) \frac{f(c)}{F(d_1)} dc + V(d_1 + d_0) \right) \\ &= \frac{\lambda R(d_1)}{r + \lambda F(d_1)} + \frac{\lambda F(d_1)}{r + \lambda F(d_1)} V(d_1 + d_0), \end{aligned}$$

where $\int_0^{d_1} (r^{-1} - c) \frac{f(c)}{F(d_1)} dc$ is the planner's payoff from an idea conditional on the event that its cost is below d_1 . Because $(r^{-1} - c)$ already accounts for future contributions of the innovation developed at τ , the continuation value $V(d_1 + d_0)$ must exclude those contributions to avoid double counting. This is why we define $V(\cdot)$ as the discounted value of future, but not existing, innovations. The dynamic programming equation is

$$V(d) = \max_{d_0, d_1 \in [0, r^{-1}]} \frac{\lambda R(d_1)}{r + \lambda F(d_1)} + \frac{\lambda F(d_1)}{r + \lambda F(d_1)} V(d_1 + d_0), \quad (3)$$

subject to the PK constraint (2).

As discussed above, we define sharing to be the allocation of rights to multiple innovators at the same history:

Definition 2 *An optimal policy has **sharing** at d if $d_0(d) > 0$ and $d_1(d) > 0$.*

We describe in Section 4 the sense in which such policies might generate cost, and the sense in which policies that do not have sharing are more easily adjudicated. For now we simply state the definition, and study whether policies involve sharing, so defined.

The optimal policy has three regions, as defined in the following proposition. The proof of the proposition is contained in Appendix A; here we state the optimal policy and explain the intuition behind it.

⁸Strictly speaking, the constraint $y \geq 0$ makes (1) more restrictive than (2). However, the constraint $y \geq 0$ never binds in our model and does not play any role.

Proposition 1 Define $\bar{d}$ to be the value in $(0, r^{-1})$ such that

$$r\bar{d} + \lambda F(\bar{d})\bar{d} = 1.$$

The optimal policy rule is:

(i) if $d > \bar{d}$, then $d_0 + d_1 = d$ and the PK constraint (2) binds, i.e.,

$$d_0(d) = d - h(1 - rd) > 0, \quad d_1(d) = h(1 - rd) > 0,$$

where $h(\cdot)$ is the inverse of $\lambda F(d_1)d_1$;⁹

(ii) there is a number $d^* \in (0, \bar{d})$, such that if $d \in [d^*, \bar{d}]$, then $d_0 = 0$ and (2) binds, i.e., $d_1(d) = F^{-1}\left(\frac{1-rd}{\lambda d}\right) \equiv g(d)$;

(iii) if $d < d^*$, then $d_0 = 0$, $d_1(d) = g(d^*)$ and (2) is slack.

Remark Both $g(d)$ and $h(1 - rd)$ are decreasing in d . Because $g(d)$ and $h(1 - rd)$ represent the optimal d_1 when $d_0 = 0$ and $d_0 > 0$ respectively, their monotonicity reflects the fundamental scarcity in this model: a greater promise to incumbents limits what can be offered to future innovators. Both $d_1(\cdot)$ and $d_0(\cdot)$ are continuous at $\bar{d}$ because $g(\bar{d}) = h(1 - r\bar{d}) = \bar{d}$ and $\bar{d} - h(1 - r\bar{d}) = 0$.

Start with $d = \bar{d}$. In this case, $d_1 = \bar{d}$ and $d_0 = 0$. So $d = \frac{1}{r + \lambda F(d_1)}$: the duration promise of $d = \bar{d} = d_1$ is exactly delivered by granting to the incumbents all instants before a new innovator whose cost is below $\bar{d}$ arrives. This uses up every instant of time, and perfectly smooths duration across innovators (including the incumbents); therefore, $\bar{d}$ is the highest promise for all innovators if the total duration of r^{-1} is divided equally. Smoothing is valuable, intuitively, because nonsmooth paths (where the cutoff $\bar{c}_t$ varies over time) substitute higher cost innovations for lower cost innovations. Therefore, starting from $\bar{d}$, it is optimal to use all instants and perfectly smooth through a constant cutoff $\bar{d}$.

⁹Because $d_0 + d_1 = d$, the binding PK constraint is $d = \frac{1}{r + \lambda F(d_1)} + \frac{\lambda F(d_1)}{r + \lambda F(d_1)}d_0 = \frac{1}{r + \lambda F(d_1)} + \frac{\lambda F(d_1)}{r + \lambda F(d_1)}(d - d_1)$, which implies $1 - rd = \lambda F(d_1)d_1$, or $d_1 = h(1 - rd)$. That $d_0(d) > 0$ is because $\bar{d} - h(1 - r\bar{d}) = 0$ and $d - h(1 - rd)$ increases in d .

This logic extends to $d > \bar{d}$, with slight modification. When $d > \bar{d}$, market time is insufficient to cover a sequence of duration promises of $d_1 = \bar{d}$. However, the planner can use up all instants, and perfectly smooth the remaining time, by offering a constant but lower duration promise to all future innovators and leaving the rest to satisfy the initial duration promise. To see this, note that the policy for $d > \bar{d}$ implies a binding PK constraint (2) (i.e., all instants are used) and $d_0 + d_1 = d$ (i.e., the total duration promise stays at d forever and the cutoff $\bar{c}_t$ equals $d_1 = h(1 - rd) < \bar{d}$ forever). Since both d_1 and d_0 are positive in this range, there is sharing.

When $d < \bar{d}$, the policy function must depart from either perfect smoothing or the use of all future instants, or both. This is because if the planner followed the policy for $d > \bar{d}$ to achieve the two goals, then $d_0 = d - h(1 - rd)$ is negative when $d < \bar{d}$. To satisfy the nonnegativity constraint, it is not surprising that the planner sets $d_0 = 0$ when $d < \bar{d}$; doing so necessarily harms at least one of the two goals.

When $d < \bar{d}$, because $d_0 = 0$, the PK constraint $d \leq \frac{1}{r + \lambda F(d_1)}$ can be rewritten as $d_1 \leq F^{-1}\left(\frac{1 - rd}{\lambda d}\right) \equiv g(d)$. Since $g(d)$ decreases in d , the PK constraint binds if and only if d exceeds some threshold, denoted as d^* . In $[d^*, \bar{d}]$, the PK constraint binds and $d_1 = g(d)$; below d^* , d_1 equals the unconstrained maximizer $g(d^*)$.

Finally, we explain why the unconstrained maximizer of d_1 is above $\bar{d}$, or equivalently, why $d^* < \bar{d}$. If $d_1 = \bar{d}$, there is a first-order benefit from raising d_1 above $\bar{d}$: the first innovator who innovates will arrive sooner. The downside is that it distorts future arrivals because $d_1 > \bar{d}$ implies that the first innovation gets more duration than all subsequent ones (i.e., $d_1 > h(1 - rd_1)$). But for d_1 slightly above $\bar{d}$ this distortion cost is not first order, because the duration is nearly perfectly smoothed. Hence the planner will choose $d_1 > \bar{d}$ when the PK constraint $d_1 \leq g(d)$ is slack.

With the policy in hand, we can discuss dynamics, and the role of sharing. Generically, the optimal policy jumps to a point where there is perpetual sharing.

Corollary 1 *For any initial d , the state variable $(d_0 + d_1)$ jumps immediately to a constant level. For initial $d \neq \bar{d}$, this constant level is strictly greater than $\bar{d}$, so d_0 and d_1 are both strictly positive, i.e., there is sharing forever.*

We generate optimal allocations where more than one innovator holds a claim to future profits. This contrasts with Hopenhayn et al. (2006), where sufficient conditions are developed such that the optimal system does not involve sharing. They assume there is sufficient heterogeneity in types to ensure that sharing is never optimal; the planner concentrates rights in the hands of a few innovators who use rights most efficiently. We offer a contrast, where heterogeneity is not as extreme, and therefore the optimal policy generates sharing. Presumably heterogeneity varies across different industries, and therefore one might think of the papers as providing guidance as to why sharing is more prevalent in some industries than in others.¹⁰ In Appendix B we discuss the relationship between heterogeneity and sharing in more detail.

Example 1 (Uniform density) *To further build understanding of the optimal policy, suppose that the density f is uniform on $[0, 1]$ and $r = \lambda = 1$. In this case we can solve for the optimal policy analytically by solving polynomial equations. Since $\bar{d}$ satisfies $\bar{d}^2 + \bar{d} = 1$, it is $\frac{-1+\sqrt{5}}{2}$, the golden ratio. The value function is*

$$V(d) = \begin{cases} (1 - d^*) (0.5 + \sqrt{2 - d^{*-1}}), & d \leq d^* \approx 0.591; \\ (1 - d) (0.5 + \sqrt{2 - d^{-1}}), & d \in [d^*, \bar{d}]; \\ \sqrt{1 - d} - 0.5(1 - d), & d \geq \bar{d}. \end{cases}$$

Figures 1 and 2 plot the value function and the policy functions. In Figure 2, d_1 has three segments. The flat part where $d < d^$ corresponds to the region where the PK constraint does not bind. The segment between d^* and $\bar{d}$ is defined by $d_1(d) = g(d) = \frac{1-d}{d}$; the last segment for $d > \bar{d}$ is described by the policies $d_0 = d - \sqrt{1 - d} > 0$ and $d_1 = \sqrt{1 - d} > 0$.*

4 Interpreting the Optimal Contract with Sharing

The previous section defines sharing to be the case where both the current and past innovator are promised some time selling the leading-edge product; the optimal contract

¹⁰Industries may differ in the degree of heterogeneity of innovation opportunities, for instance because some industries are more characterized by exploration and others by exploitation (see Akcigit and Kerr (2010)).

Figure 1: Value function with uniform density.

Figure 2: Policy functions with uniform density.

employs such sharing, from (at the latest) the second innovation onward. We describe here several interpretations of such a policy.

One interpretation of the optimal contract is as the design of an optimal policy for patents. We see, on the one hand, that some ideas would not be allowed to profit at all; ideas with cost greater than d_1 are not offered sufficient protection to be developed. We interpret this as being unpatentable, although our mechanism design approach implies that this decision is made completely by the innovator and is not adjudicated by the planner; the planner simply sorts out what she seeks to be unpatentable by a sufficiently high patent fee.

On the other hand, innovations that are allowed to generate profits for the innovator (i.e., patents issued) share with prior patents. In practice, this could occur in several ways. For instance, when a new patent infringes on an old one, the two parties must come to a licensing agreement. Alternatively, if infringement is not clear, the firms could potentially engage in litigation. We argue that both outcomes can be interpreted as part of the optimal contract with sharing.

One preset rule that allocates rights with sharing is a lottery. Suppose there is one incumbent with a promise, d . When sharing is called for, the new innovator and the incumbent are prescribed d_1 and d_0 , respectively. Rather than maintain that promise for both firms, the planner could have a lottery: one of the two will be chosen to be the new incumbent, and given a promise of $d_0 + d_1$. A lottery assigns the identity of the new incumbent, with, for instance, the new innovator replacing the old incumbent with probability $\frac{d_1}{d_0 + d_1}$. Because the innovators are risk neutral, this lottery version of sharing achieves the same allocation. In the tradition of the literature on weak patent rights, a natural interpretation of probabilistic protection is as litigation. Here the policy uses litigation as a method of allocating profits to different contributors; the odds of a firm winning the litigation is tied to its share of the duration promise offered to the two innovators under the optimal policy.

Alternatively, one can interpret sharing as coming through licensing contracts under the circumstance where later patents infringe on early ones. One can interpret the licensing rules as part of the social planner's patent policy, or as a set of licensing

agreements arranged ex ante among the potential innovators; that is, the patent policy offers a right of exclusion to all the rights holders who have a share of the profits, and the rights holders have agreed to a preset sharing rule at time zero that maximizes expected surplus of all potential innovators.

Under the licensing interpretation, the innovators form an ever-increasing pool of the patents that have arrived. If the first innovator has full commitment power, then he can be granted a broad patent to which he commits to follow the optimal allocation with regard to future innovations, sharing rights in exchange for fees from future innovators. Under that interpretation the model is in the spirit of Green and Scotchmer (1995). The allocation is also similar to standard notions of a patent pool in the sense that many innovators have jointly contributed to the research line, and as such, they would share in the profits of their joint effort. Here, however, the pool is ever growing as a result of the ever-improving nature of the product. More specifically, new innovators may join the pool for a fee; in exchange they receive a fraction $\frac{d_1}{d_0+d_1}$ of pool profits. Whenever a new innovator joins the pool, then, all the existing innovators' shares drop by the same proportion to make room. A new innovator's share is initially $\frac{d_1}{d_0+d_1}$; it is reduced to $\frac{d_1}{d_0+d_1}(1 - \frac{d_1}{d_0+d_1})$ after the second innovator joins, and further reduced to $\frac{d_1}{d_0+d_1}(1 - \frac{d_1}{d_0+d_1})^2$ after the third innovator joins, etc. One can verify that the expected discounted shares for each new innovator is indeed d_1 , the duration promise in the planner's allocation.¹¹

In the setup we have used, the existing innovators would never want to exclude a new innovator willing to pay the entry fee, even if they weren't obligated to by the commitment of the contract. To see this, notice that the marginal innovator $c = d_1$ makes zero profit from joining the pool and improves total pool profits by contributing an improvement; lower cost types contribute the same. However, if contracting could only take place after the new improvement spent c , there would be a simple hold-up

¹¹Denote the arrival time of the n th innovator whose cost is below d_1 as τ_n . The first innovator's share is $\frac{d_1}{d_0+d_1}(1 - \frac{d_1}{d_0+d_1})^{n-1}$ between τ_n and τ_{n+1} , and the expected discounted (to τ_1) duration of the random interval $[\tau_n, \tau_{n+1}]$ is $\frac{1}{r+\lambda F(d_1)} \left(\frac{\lambda F(d_1)}{r+\lambda F(d_1)} \right)^{n-1}$. Therefore, the first innovator's expected discounted shares is $\sum_{n=1}^{\infty} \frac{d_1}{d_0+d_1} (1 - \frac{d_1}{d_0+d_1})^{n-1} \frac{1}{r+\lambda F(d_1)} \left(\frac{\lambda F(d_1)}{r+\lambda F(d_1)} \right)^{n-1} = \frac{\frac{d_1}{d_0+d_1} \frac{r+\lambda F(d_1)}{r+\lambda F(d_1)}}{1 - \frac{\lambda F(d_1)}{r+\lambda F(d_1)} \frac{d_0}{d_0+d_1}} = \frac{\frac{r+\lambda F(d_1)}{r+\lambda F(d_1)} d_1}{d_0+d_1 - \frac{\lambda F(d_1)}{r+\lambda F(d_1)} d_0} = d_1$, where the last equality follows from the binding PK constraint $d_0 + d_1 = \frac{1}{r+\lambda F(d_1)} + \frac{\lambda F(d_1)}{r+\lambda F(d_1)} d_0$.

problem. One can interpret the planner’s role here as to limit this hold-up. To do so, the planner should insist on a “non-discriminatory” policy for new pool entrants that forces the pool to pre-specify the “fair” or “reasonable” price at which they will allow new members to join the pool, and accept membership from anyone who wants to pay the entry fee.

This gives a new role for regulation of patent pools. Policymakers have insisted that pools treat users of the pool’s patents in a way that is FRAND.¹² The motivation for this policy is that a patent pool among a fixed set of innovators is like a merger between the members, and therefore care must be exercised to make sure that patent pools don’t have the anticompetitive effects of mergers on the pool’s users.¹³ The model proposed here considers how pools should be allowed to contract with potential new members, given the fact that new members increase total profits but erode pool members’ share of the profits. To focus on the issue of how pools form, we specifically study a case where there are no welfare consequences of the pool’s treatment of the users of the pool’s product.

Note that policies without sharing can be implemented in a much simpler way from the licensing or litigation interpretations. Consider the optimal policy starting from $d = \bar{d}$. In this case the policy can be decentralized through a rule that depends only on reports of arrivals. In particular, each arrival needs only to pay an entry fee, at which point they are given the sole right to profit; that is, a completely exclusionary patent that infringes on nothing, and allows the holder to exclude all past innovators. The innovator who most recently paid the fee unambiguously has all the rights.

Consider, by contrast, some initial $d > \bar{d}$ where there is forever $d_0 > 0$ and $d_1 > 0$. Here decentralization requires something other than just reports of arrivals; sharing rules conditional on those reports are essential. We view all of these constructions as potentially generating costs relative to cases without sharing. In the next section we consider the extreme case, where sharing is so costly that the planner must avoid it altogether.

¹²See, for example, Lerner and Tirole (2008).

¹³This idea is the basis of the model of patent pools in Lerner and Tirole (2004), where pools have welfare consequences similar to the ones found in models of mergers, based on monopoly markup by the pool.

5 Conflict-Free Policies

Optimal policies in Section 3 included a particular sense of potential conflict, stemming from sharing between multiple rights holders promised at a given history. The planner could avoid this sort of conflict by restricting attention to policies without sharing, which, upon the arrival of a new innovation, ends the rights of previous rights holders. This translates to the same model of Section 3, but under the restriction that $d_0 = 0$. Since avoiding conflict in this sense adds a constraint, doing so always comes at a cost. In this section we will ask two questions: First, what are the implications of following such a policy? And second, what can we say about the costs imposed by avoiding conflict?

5.1 Optimal Policies

Recall that it is optimal to choose $d_0 = 0$ in the region $[0, \bar{d}]$ in Proposition 1. There the PK constraint $d_1 \leq F^{-1}\left(\frac{1-rd}{\lambda d}\right) \equiv g(d)$ binds if and only if d exceeds a threshold d^* . Here the PK constraint is also $d_1 \leq g(d)$ because $d_0 = 0$ is imposed. Unsurprisingly, the PK constraint binds if and only if d exceeds some threshold, denoted as d^{**} . Above d^{**} , $d_1 = g(d)$; below d^{**} , d_1 equals the unconstrained maximizer $g(d^{**})$.

Proposition 2 *When $d_0 = 0$ is imposed in (2) and (3), the optimal policy rule is*

$$d_1(d) = \begin{cases} g(d^{**}), & d \leq d^{**}; \\ g(d), & d \geq d^{**}, \end{cases}$$

where d^{**} is a number in $(0, \bar{d}]$. The rule can be summarized as $d_1 = \min(g(d), g(d^{**}))$.

The evolution of duration promises critically depends on whether d^{**} equals $\bar{d}$. The next proposition provides an easy-to-verify condition to check this.

Proposition 3 *$d^{**} = \bar{d}$ if and only if $F(\bar{d}) \geq \bar{d}f(\bar{d})$.*

If the density function is weakly decreasing, then $d^{**} = \bar{d}$; if the density function is strictly increasing, then $d^{**} < \bar{d}$.

The condition for $d^{**} = \bar{d}$ is that $f(\bar{d})$ is bounded above by $F(\bar{d})\bar{d}^{-1}$. To understand why the density cannot be too large in this case, consider the optimal path of duration promises. Denote the duration promise of the n th innovation as d_n (i.e., d_n satisfies $d_n = d_1(d_{n-1})$). When $d^{**} = \bar{d}$, the optimal path is perfectly smoothed, i.e., $d_n = \bar{d}$ for all n . Suppose we deviate and increase d_1 above $\bar{d}$. Because $d_2 = \min(g(d_1), g(d^{**})) = \min(g(d_1), \bar{d}) = g(d_1)$, a higher d_1 reduces d_2 below $\bar{d}$. Note that this deviation does not affect d_n for $n \geq 3$. Hence the benefit of a higher d_1 must be weighted against the cost of a lower d_2 . The cost will dominate the benefit when $g(d_1)$ is sensitive in d_1 . The derivative of $g(\cdot)$ is proportional to the reciprocal of the density. Hence a lower density at $\bar{d}$ implies a more sensitive $g(\cdot)$, which further implies that cost dominates benefit and $d_1 = \bar{d}$ is optimal.

In the case with sharing the unconstrained optimal d_1 is strictly above $\bar{d}$, since the cost of raising d_1 above $\bar{d}$ is not first order: small increases in d_1 above $\bar{d}$ are perfectly smoothed across future ideas through sharing. Without sharing, however, any increase in d_1 above $\bar{d}$ cannot be smoothed, and therefore there is a first-order cost in raising d_1 above $\bar{d}$. This implies that it is possible in this case to have the unconstrained optimal $d_1 = \bar{d}$ (i.e., $d^{**} = \bar{d}$).

5.2 Dynamics without Sharing

When $d^{**} = \bar{d}$, the dynamics of duration promises are simple: if $d \leq \bar{d}$, then $d_1 = \bar{d}$; otherwise if $d > \bar{d}$, then $d_1 < \bar{d}$ and $d_2 = \bar{d}$. Hence we have

Corollary 2 *If $d^{**} = \bar{d}$, then the state variable reaches $\bar{d}$ in at most two innovations.*

Figure 3 plots the conflict-free policy function with uniform density function. It has two segments. The segment where $d \leq \bar{d}$ is where the PK constraint does not bind, while the segment for $d > \bar{d}$ is defined by $d_1(d) = g(d)$.

When $d^{**} < \bar{d}$, the distinctive feature of the dynamics of duration promises is that they cycle. This is the clear sense in which a conflict-free policy always has more variable technological progress, which we discuss in more detail below.

Figure 3: Conflict-free policy function with uniform density.

Proposition 4 *If $d^{**} < \bar{d}$, then either the promises to all odd innovations or all even innovations are above $\bar{d}$. Without loss of generality, suppose $d_n \geq \bar{d}$ for all odd n . Then there exists a $d_\infty \geq \bar{d}$ such that $\lim_{n \rightarrow \infty} d_{2n+1} = d_\infty$ and $\lim_{n \rightarrow \infty} d_{2n} = g(d_\infty)$.*

When $d^{**} < \bar{d}$, the states fluctuate around $\bar{d}$ because if $d_n > \bar{d}$, then $d_{n+1} = g(d_n) < \bar{d}$. This further implies that $d_{n+2} = \min(g(d_{n+1}), g(d^{**})) > \bar{d}$, as both $g(d_{n+1})$ and $g(d^{**})$ are above $\bar{d}$. Intuitively, when the current patent protection is large, the planner cannot implement many innovations, and therefore offers a small reward to potential innovators. Once an innovator accepts the reward, the planner no longer has to deal with such a large patent in place, and therefore can promise more generous duration to the subsequent innovation. This generous duration promise, once offered to a subsequent innovator, brings the situation back to a high level of protection.

The two-period cycle in Proposition 4 can either be forever fluctuating (i.e., $d_\infty > \bar{d}$ and $g(d_\infty) < \bar{d}$) or converging to $\bar{d}$ (i.e., $d_\infty = g(d_\infty) = \bar{d}$). Figures 4 and 5 plot two dynamics of duration promises starting with $d_1 = 0.76$. We provide sufficient conditions for the two cases.

Proposition 5 *Suppose $d^{**} < \bar{d}$ and the initial duration promise $d \neq \bar{d}$.*

(i) *If $g(g(c)) > c$ for all $c \in (\bar{d}, r^{-1}]$, then $d_\infty = g(d^{**})$.*

(ii) *If $f(c) \leq r^2 \lambda^{-1}$ for all $c \in [0, r^{-1}]$, then $g(g(c)) > c$ for all $c \in (\bar{d}, r^{-1}]$.*

Proposition 6 *Let $b \in (\bar{d}, r^{-1})$ be the unique solution of*

$$\lambda \int_0^b (b - c) f(c) dc + rb - 1 - \lambda R(h(1 - rb)) = 0. \quad (4)$$

(i) *If $g(g(c)) < c$ for all $c \in (\bar{d}, b]$, then $d_\infty = \bar{d}$.*

(ii) *If $f(c) = r^\alpha \alpha c^{\alpha-1}$ for $c \in [0, r^{-1}]$ and α is sufficiently large, then $g(g(c)) < c$ for all $c \in (\bar{d}, b]$.*

We end this section with numerical examples with power function density.

Example 2 (Power function density) *Suppose the density is $f(c) = \alpha c^{\alpha-1}$ for $\alpha > 1$ and $r = \lambda = 1$. Figure 6 plots the policy function when $\alpha = 3$: $d_1(d_1(\cdot))$ is steeper than the 45-degree line in a neighborhood of $\bar{d}$ and cycles are amplified over time. The sufficient condition in Proposition 5 is satisfied and cycles last forever. However, when α is sufficiently large, the sufficient condition in Proposition 6 is satisfied and the states converge to $\bar{d}$. Figure 7 plots the policy function when $\alpha = 6$: in contrast to that in Figure 6, $d_1(d_1(\cdot))$ is flatter than the 45-degree line and cycles disappear eventually.*

6 Discussion

6.1 Comparison of Technological Progress With and Without Sharing

Although welfare must be (weakly) lower without sharing, the impact on the rate of innovations is somewhat more complicated. First, we consider the case of $d^{**} = \bar{d}$; in that case, there are no cycles in the conflict-free policy.

Figure 4: Perpetual fluctuation with power density $f(c) = 3c^2$.

Figure 5: Degeneration with power density $f(c) = 6c^5$.

Figure 6: Policy function with power density $f(c) = 3c^2$.

Figure 7: Policy function with power density $f(c) = 6c^5$.

6.1.1 The Case of $d^{**} = \bar{d}$

For any given d , the rate of innovation is weakly higher with sharing; it is strictly higher everywhere except $\bar{d}$.

Corollary 3 $d_1(d)$ is weakly higher with sharing, and strictly if $d \neq \bar{d}$.

This does not, however, imply that the long-run rate of progress is higher with sharing; the evolution of d_n is endogenous, and since the planner with access to sharing is giving out more promises starting from any d , that leads to more constraints later on. Define $d_{ss}(d)$ to be the steady state value of duration starting from d ; in both cases it is achieved after at most two improvements. Since steady state duration is $\bar{d}$ without sharing, and greater than $\bar{d}$ with sharing so long as initial duration is not $\bar{d}$, we have

Corollary 4 $d_1(d_{ss}(d))$ is weakly higher without sharing, and strictly if $d \neq \bar{d}$.

For any starting duration where sharing and non-sharing differ ($d \neq \bar{d}$), the same pattern emerges: sharing leads to faster progress initially, but, as a result of the higher promises given out, the long-run progress is slower. The intuition is that, with sharing, the planner perfectly smooths the duration at her disposal, offering equal duration to every arrival. The planner prefers smooth paths because they don't bypass low-cost ideas at some points in time and implement higher-cost ideas later on. Without sharing, the planner can never achieve this smoothing if the promise rises above $\bar{d}$. To deliver the large duration promise, only the lowest-cost follow-up innovations are allowed; that is, d_1 is very low. Once a follow-up innovation is developed, however, we implement further follow-ups at a constant rate $\lambda F(\bar{d})$. The welfare benefits of sharing come from the benefits of smoothing: smoother progress under sharing more efficiently implements ideas by not bypassing low-cost ideas when d_1 is low.

6.1.2 The Case of $d^{**} < \bar{d}$

When $d^{**} < \bar{d}$, no sharing clearly leads to more variable growth, since promises cycle. In terms of the short-run rate of progress, if $d \geq \bar{d}$, then $d_1(d)$ is higher with sharing,

and therefore conflict-free policies have slower technological progress. If $d < \bar{d}$, whether $d_1(d)$ is higher with sharing depends on whether $d^* < d^{**}$.¹⁴

The comparison of long-run growth with and without sharing could go either way in this case. To see why the lack of sharing might lead to slower long-run growth, consider example 2, where the contract enters a perpetual cycle for all $d > \bar{d}$. The long-run growth rate fluctuates between $\lambda F(g(d_\infty))$ and $\lambda F(g(g(d_\infty)))$, and the average may be lower than $\lambda F(\bar{d})$. If the initial promise is only slightly above $\bar{d}$, the growth rate of the sharing contract will be arbitrarily close to $\lambda F(\bar{d})$, and dominate the average growth rate without sharing.

In the next subsection, we further the point about sharing and convexification by describing an environment where the initial duration promise may be large, due to the special costs that might come with being a market pioneer. In that environment, the feature that non-sharing leads to faster long-run growth is restored, as the growth rate of the sharing contract is sufficiently low. In that environment, the feature in Corollary 4 that non-sharing leads to faster long-run growth is restored.

6.2 Application: Ironclad Patents and Rewarding a Market Pioneer

A fundamental force in the model is that rewards for innovation come through market profits, and those profits are limited. As such, it is natural to consider how the planner might respond to a special innovation that opens the door for future improvements but requires extra rewards in order to be developed, such as a pioneering innovation that begins the process. Our construction of the optimal contract allows for direct analysis of a first innovation that differs from subsequent innovations. Suppose that this first arrival is similar to the others, in the sense that its arrival and investment are unobserved, but is different from subsequent innovations in terms of the cost of investment and the quality improvement. To make the analysis as simple as possible, we assume, in the language of Scotchmer (1999), that the pioneering innovation has quality q and a deterministic

¹⁴Although we could not prove $d^* < d^{**}$ analytically, it holds in all of our numerical examples.

cost c . Since the pioneer's benefit from innovation is $d_p q$ when he is allocated discounted duration d_p , pioneering only occurs if the pioneer is offered $d_p \geq c/q$. In other words, if either the cost of pioneering is unusually large (as documented in Robinson et al. (1994)) or the initial quality is low (as is natural if the pioneering innovation is mostly valued for its ability to generate more marketable improvements), the pioneer must be offered a high duration promise. The system then evolves as in prior sections, with d_p acting as an initial condition for the duration promise.

We focus on the case where the promise d_p to the pioneer satisfies $g(g(d_p)) = r^{-1}$. The promise is large in the sense that the PK constraint binds with and without sharing. If sharing is possible, a high promise to the pioneer is realized by continuous sharing: the rate of innovation is constant, and that rate is lower the larger is the pioneer's promise. A larger promise translates to a greater share of future profits for the pioneer, and therefore only lower-cost improvements are profitable.

If sharing is impossible, an initial large promise d_p leads to an initial rate of progress lower than with sharing, as the high duration promise to the pioneer can only be realized through severe exclusion restrictions. One can think of such a patent as ironclad in the sense that it keeps many potential entrants out of the market. However, once an idea whose cost is lower than $g(d_p) = (r + \lambda)^{-1}$ arrives, it will be developed and break the ironclad patent. In other words, the pioneer's rights are stronger without sharing, but are fully gone sooner. Since the duration promise to the low-cost idea, $g(d_p)$, equals $(r + \lambda)^{-1}$, PK constraint becomes slack immediately after the development of the low-cost idea. The continuation contract starts afresh as if the planner is not committed to any duration promise. As we mentioned, sharing leads to smooth progress due to constant sharing with the pioneer, while avoiding conflict forces the planner to temporarily reduce progress, but allows progress to rise later on. This leads to a less smooth path of progress without sharing, which is costly to the planner.

7 Conclusion

In this paper we have constructed optimal allocations for a sequence of innovators who, due to moral hazard, must be rewarded with profit-making opportunities. We have shown that the optimal allocations involve sharing so that more than one firm get a share of future profits. We interpret this sharing as patents that infringe on prior art, together with licensing. We show how the licensing contract can be interpreted as an ever-growing patent pool and provide theoretical foundations for observed practices like patentability requirements and infringement, as well as weak patent rights. By constructing allocations that do not allow the planner to use shared rewards, we can explore the role of licensing contracts in technological progress. Sharing contracts lead to smoother progress. They also lead to faster progress initially.

We focus on the extreme case where the planner either uses sharing, or the cost of sharing is infinite. A natural topic for future research is to see what degree of sharing the planner would choose if faced with a finite cost of assigning shared rights. The trade-off in making that decision is highlighted by the analysis here: Sharing is valuable as a convexification device.

Appendix A: Proofs

For convenience, we rewrite the Bellman equation (3) into the following (5). Define $\tilde{d} \equiv \frac{1}{r + \lambda F(d_1)}$, and hence $1 - r\tilde{d} = \frac{\lambda F(d_1)}{r + \lambda F(d_1)}$. Using $g(\tilde{d}) \equiv F^{-1}\left(\frac{1 - r\tilde{d}}{\lambda \tilde{d}}\right) = d_1$, we can rewrite (3) as

$$\begin{aligned} V(d) &= \max_{\tilde{d}, d_0 \in [0, r^{-1}]} \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V(g(\tilde{d}) + d_0) \\ &\text{subject to} \quad d \leq \tilde{d} + (1 - r\tilde{d})d_0. \end{aligned} \tag{5}$$

Proposition 1 *When $d_0 \geq 0$ is imposed, the solution to (5) is*

$$V(d) = \begin{cases} d^*\lambda R(g(d^*)) + (1 - rd^*)V_{\mathbb{R}}(g(d^*)), & d \leq d^*; \\ d\lambda R(g(d)) + (1 - rd)V_{\mathbb{R}}(g(d)), & d \in [d^*, \bar{d}]; \\ V_{\mathbb{R}}(d) \equiv r^{-1}\lambda R(h(1 - rd)), & d \geq \bar{d}, \end{cases} \tag{6}$$

where $V_{\mathbb{R}}(\cdot)$ is the relaxed value function defined in Lemma A.3 and d^* is the unconstrained maximizer of $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}))$. $V(\cdot)$ is concave and its derivative

is continuous at $\bar{d}$. The optimal policy rule is

$$(d_0(d), d_1(d)) = \begin{cases} (0, g(d^*)), & d \leq d^*; \\ (0, g(d)), & d \in [d^*, \bar{d}]; \\ (d - h(1 - rd), h(1 - rd)), & d \geq \bar{d}. \end{cases}$$

Proof: This proof proceeds in several steps, and relies on the property of $V_{\mathbb{R}}(\cdot)$ in Lemma A.3.

First, we show the concavity of $V(\cdot)$ both below and above $\bar{d}$. $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}))$ is concave in $\tilde{d}$, because Lemmas A.1, A.2 and A.3 show that $\tilde{d}R(g(\tilde{d}))$ and $(1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}))$ are concave in $\tilde{d}$, and $V_{\mathbb{R}}(d)$ is concave in d .

Second, we show that the maximizer $d^* < \bar{d}$. It is sufficient to verify that

$$\left(\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d})) \right)' \Big|_{\tilde{d}=\bar{d}} < 0.$$

It follows from the definition of $V_{\mathbb{R}}(\cdot)$ that

$$\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d})) \leq V_{\mathbb{R}}(\tilde{d}), \quad \forall \tilde{d},$$

and the equality holds at $\tilde{d} = \bar{d}$. Because both $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}))$ and $V_{\mathbb{R}}(\tilde{d})$ are concave, they must be tangent at $\bar{d}$, that is

$$\left(\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d})) \right)' \Big|_{\tilde{d}=\bar{d}} = V'_{\mathbb{R}}(\bar{d}) < 0. \quad (7)$$

Third, $V'(\cdot)$ is continuous at $\bar{d}$, which implies that $V(\cdot)$ is globally concave. It follows from (7) that

$$\lim_{d \uparrow \bar{d}} V'(d) = V'_{\mathbb{R}}(\bar{d}) = \lim_{d \downarrow \bar{d}} V'(d) = \lim_{d \downarrow \bar{d}} V'(d).$$

Fourth, we verify the Bellman equation (5). Because V coincides with $V_{\mathbb{R}}$ when $d \geq \bar{d}$, we only verify (5) on $[0, \bar{d}]$. Pick a feasible $(d_0, \bar{d})$ such that $d_0 > 0$. We will show that $V(d) \geq \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}) + d_0)$ as follows. If $\tilde{d} \geq \bar{d}$, then

$$\begin{aligned} \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}) + d_0) &< \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d})) \\ &\leq \bar{d}\lambda R(g(\bar{d})) + (1 - r\bar{d})V_{\mathbb{R}}(g(\bar{d})) \\ &= V(\bar{d}) \leq V(d), \end{aligned}$$

where the second inequality follows from $\left(\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d})) \right)' \Big|_{\tilde{d}=\bar{d}} < 0$. If $\tilde{d} < \bar{d}$, then

$$\begin{aligned} &\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}) + d_0) \\ &= \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d})) + (1 - r\tilde{d}) \int_0^{d_0} V'(g(\tilde{d}) + x) dx \\ &\leq V(\tilde{d}) + (1 - r\tilde{d}) \int_0^{d_0} V'(\tilde{d} + (1 - r\tilde{d})x) dx = V(\tilde{d} + (1 - r\tilde{d})d_0) \leq V(d), \end{aligned}$$

where the first inequality relies on $g(\tilde{d}) + x \geq \tilde{d} + (1 - r\tilde{d})x$ and the monotonicity of V' , and the second inequality relies on the monotonicity of $V(\cdot)$.

Finally, the policy rule can be easily derived from the value function $V(\cdot)$. $\blacksquare$

When $d_0 = 0$ is imposed in (5), the Bellman equation becomes

$$W(d) = \max_{\tilde{d}} \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})W(g(\tilde{d})), \quad \text{subject to } d \leq \tilde{d},$$

where $W(\cdot)$ denotes the conflict-free value function.

Proof of Proposition 2: First, we show that W is weakly decreasing and concave. Let $B([0, r^{-1}])$ be the collection of bounded functions on $[0, r^{-1}]$ and define an operator $T : B([0, r^{-1}]) \rightarrow B([0, r^{-1}])$ by

$$(Tw)(d) = \max_{\tilde{d} \in [(r+\lambda)^{-1}, r^{-1}]} \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})w(g(\tilde{d})), \quad \text{subject to } d \leq \tilde{d}.$$

We can easily verify that T satisfies the Blackwell's sufficient conditions and is a contraction mapping. Hence T has a unique fixed point $W(\cdot)$. To show that $W(\cdot)$ is weakly decreasing and concave, it is sufficient to prove that T maps any weakly decreasing and concave function $w(\cdot)$ into a weakly decreasing and concave function. It follows from Lemmas A.1 and A.2 that $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})w(g(\tilde{d}))$ is concave in $\tilde{d}$. Therefore, $(Tw)(\cdot)$ is concave. The monotonicity of $(Tw)(\cdot)$ follows from the fact that the feasibility set in the Bellman equation shrinks with higher d .

Second, we verify the optimal policy rule for d_1 . Let d^{**} be the unique maximizer of $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})W(g(\tilde{d}))$. Then $W(\cdot)$ is flat below d^{**} , but strictly decreasing above d^{**} . In other words, the PK constraint $d \leq \tilde{d}$ binds if and only if $d > d^{**}$. When it binds, the choice of d_1 is pinned down by the constraint $\tilde{d} = d$, that is, $d_1 = g(d)$. When the PK constraint doesn't bind, $\tilde{d} = d^{**}$ and $d_1 = g(d^{**})$. To summarize, the optimal policy is $\tilde{d}(d) = \max(d, d^{**})$, or equivalently $d_1(d) = \min(g(d), g(d^{**}))$.

Third, we show that $d^{**} \leq \bar{d}$. By contradiction, suppose $d^{**} > \bar{d}$, then $g(d^{**}) < \bar{d} < d^{**}$. Hence, $W(g(d^{**})) = W(d^{**})$ and the Bellman equation implies

$$W(d^{**}) = d^{**}\lambda R(g(d^{**})) + (1 - rd^{**})W(g(d^{**})) = d^{**}\lambda R(g(d^{**})) + (1 - rd^{**})W(d^{**}),$$

which implies

$$W(d^{**}) = r^{-1}\lambda R(g(d^{**})) < r^{-1}\lambda R(\bar{d}).$$

This is a contradiction as the planner can obtain at least $r^{-1}\lambda R(\bar{d})$ by choosing $\tilde{d} = d_1 = \bar{d}$ and keeping duration promise at $\bar{d}$ forever. $\blacksquare$

Proof of Proposition 3: Necessity: If $d^{**} = \bar{d}$, then

$$W(d) = \begin{cases} r^{-1}\lambda R(\bar{d}), & d \leq \bar{d}; \\ d\lambda R(g(d)) + (1 - rd)W(\bar{d}), & d > \bar{d}. \end{cases} \quad (8)$$

If $d \leq \bar{d}$, then $g(d) \geq \bar{d}$. Then $d = \bar{d}$ solves

$$\max_{d \leq \bar{d}} d\lambda R(g(d)) + (1 - rd)W(g(d)) = \max_{d \leq \bar{d}} d\lambda R(g(d)) + (1 - rd)\Gamma(g(d)) = \max_{d \leq \bar{d}} \Omega(d),$$

where $\Gamma(\cdot)$ and $\Omega(\cdot)$ are defined in Lemmas A.4 and A.5. Hence the first-order condition $\Omega'(d)|_{d=\bar{d}} \geq 0$ and Lemma A.5 imply that $F(\bar{d}) \geq \bar{d}f(\bar{d})$.

Sufficiency: If $F(\bar{d}) \geq \bar{d}f(\bar{d})$, then we show that $W(d)$ defined in (8) satisfies the Bellman equation $W = TW$.

When $d > \bar{d}$, to show that $\tilde{d} = d$ is optimal, we need to show that $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})W(\tilde{d})$ decreases in $\tilde{d} \in [\bar{d}, r^{-1}]$. Because $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})W(\tilde{d})$ is concave in $\tilde{d}$, the monotonicity follows from

$$\left(\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})W(\tilde{d}) \right)' \Big|_{\tilde{d}=\bar{d}} = \Gamma'(\tilde{d}) \Big|_{\tilde{d}=\bar{d}} < 0,$$

where the inequality is shown in Lemma A.4.

When $d < \bar{d}$, we need to show that $\tilde{d} = \bar{d}$ is optimal. First, if $\tilde{d} > \bar{d}$, then the value achieved is $W(\tilde{d}) < W(\bar{d})$, hence $\tilde{d} > \bar{d}$ is not optimal. Second, if $\tilde{d} < \bar{d}$, then we need to show that the value achieved, $\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})W(g(\tilde{d})) = \Omega(\tilde{d})$, increases in $\tilde{d} \in [0, \bar{d}]$. It follows from Lemma A.5 and $F(\bar{d}) \geq \bar{d}f(\bar{d})$ that $\Omega'(d)|_{d=\bar{d}} \geq 0$. Hence the monotonicity of $\Omega(\cdot)$ on $[0, \bar{d}]$ follows from concavity and $\Omega'(d)|_{d=\bar{d}} \geq 0$. ■

Proof of Proposition 4: Suppose $d_n \geq \bar{d}$ for all odd n . To show that a bounded sequence $\{d_{2n+1}; n \geq 0\}$ converges, it suffices to show that it is monotone. If $d_1 \leq d_3$, then because $g(\cdot)$ is decreasing, $d_2 = g(d_1) \geq g(d_3) = d_4$, which implies

$$d_3 = \min(g(d_2), g(d^{**})) \leq \min(g(d_4), g(d^{**})) = d_5.$$

By induction, the sequence $\{d_{2n+1}; n \geq 0\}$ is increasing in n . A symmetric argument shows that the sequence is decreasing in n if $d_1 \geq d_3$. ■

Proof of Proposition 5:

- (i) If $d \in (\bar{d}, g(d^{**}))$, then $d_1 < \bar{d}$ and $d_2 = g(g(d)) > d$. Hence $\{g^{2n}(d); n \geq 0\}$ strictly increases with n , where g^{2n} denotes the composition of the function g with itself $2n$ times. This means that $g^{2n}(d)$ exceeds $g(d^{**})$ in finite time. Specifically, suppose $g^{2\bar{n}}(d) \leq g(d^{**})$ and $g^{2\bar{n}+2}(d) > g(d^{**})$ for some $\bar{n}$. Then

$$\begin{aligned} d_{2\bar{n}+1} &= g(d_{2\bar{n}}) = g(g^{2\bar{n}}(d)) < \bar{d}, \\ d_{2\bar{n}+2} &= \min(g(d_{2\bar{n}+1}), g(d^{**})) = g(d^{**}), \\ d_{2\bar{n}+3} &= g(g(d^{**})), \\ d_{2\bar{n}+4} &= \min(g(d_{2\bar{n}+3}), g(d^{**})) = g(d^{**}). \end{aligned}$$

That is, the states will cycle between $g(d^{**})$ and $g(g(d^{**}))$ starting from $2\bar{n} + 2$.

- (ii) If $f(c) \leq \lambda^{-1}r^2$ for all $c \in [0, r^{-1}]$, then $g'(d) = -\frac{(r+\lambda F(g(d)))^2}{\lambda f(g(d))} < -1$ for all d , and $(g(g(d)))' > 1$ for all d . If $d > \bar{d}$, then the mean value theorem implies $g(g(d)) - \bar{d} = g(g(d)) - g(g(\bar{d})) > d - \bar{d}$. That is, $g(g(d)) > d$ for all $d > \bar{d}$.

■

Proof of Proposition 6: We have $b \in (\bar{d}, r^{-1})$ because

$$\begin{aligned} \lambda \int_0^{\bar{d}} (\bar{d} - c)f(c)dc + r\bar{d} - 1 - \lambda R(h(1 - r\bar{d})) &< \lambda \int_0^{\bar{d}} (\bar{d} - c)f(c)dc - \lambda R(\bar{d}) < 0, \\ \lambda \int_0^{r^{-1}} (r^{-1} - c)f(c)dc + 1 - 1 - \lambda R(h(1 - 1)) &= \lambda \int_0^{r^{-1}} (r^{-1} - c)f(c)dc > 0. \end{aligned}$$

- (i) First, we show that $g(d^{**}) < b$. The first-order condition for d^{**} , $\lambda(d^{**}R(g(d^{**})))' - rW(g(d^{**})) + (1 - rd^{**})(W(g(d^{**})))' = 0$, and $W(\cdot) \leq V(\cdot)$ imply

$$\begin{aligned} \lambda(d^{**}R(g(d^{**})))' - rV(g(d^{**})) &\leq \lambda(d^{**}R(g(d^{**})))' - rW(g(d^{**})) \\ &\leq 0 = \lambda(\tilde{d}R(g(\tilde{d})))' - rV(g(\tilde{d})), \end{aligned}$$

where $\tilde{d} = \frac{1}{r + \lambda F(b)}$ and the equality follows from the definition of b in (4). That $\lambda(dR(g(d)))' - rV(g(d))$ decreases in d implies $d^{**} > \tilde{d}$, which is $g(d^{**}) < b$.

Second, $g(g(c)) < c$ for all $c \in (\bar{d}, b]$ and $g(d^{**}) < b$ imply that $g(g(c)) < c$ for all $c \in (\bar{d}, g(d^{**})]$. If $d_n \in (\bar{d}, g(d^{**})]$, then $d_{n+2} = \min(g(g(d_n)), g(d^{**})) = g(g(d_n)) < d_n$ and $\{d_{n+2m}; m \geq 0\}$ is a monotonically decreasing sequence. Similar to the proof in Proposition 5, we can show that $\lim_{m \rightarrow \infty} d_{n+2m} = \bar{d}$. Hence $d_\infty = \bar{d}$.

- (ii) If $f(c) = r^\alpha \alpha c^{\alpha-1}$, then $F(c) = r^\alpha c^\alpha$ and $R(d) = r^{\alpha-1} d^\alpha - \frac{\alpha}{\alpha+1} r^\alpha d^{\alpha+1}$. The rest of this proof contains four steps.

First, if $f(c) = r^\alpha \alpha c^{\alpha-1}$ and $x > 0$ is a small number satisfying $\frac{\lambda}{r} e^{-x} - x > 0$, then $rb < 1 - \frac{x}{\alpha}$ when α is sufficiently large. The equation for b is

$$\begin{aligned} 0 &= \lambda \int_0^b (b - c)f(c)dc + rb - 1 - \lambda R(h(1 - rb)) \\ &= \frac{\lambda (rb)^{\alpha+1}}{r^{\alpha+1}} - \frac{2\alpha + 1}{\alpha + 1} (1 - rb) + \left(\frac{\lambda}{r}\right)^{\frac{1}{\alpha+1}} (1 - rb)^{\frac{\alpha}{\alpha+1}}. \end{aligned}$$

Then $rb < 1 - \frac{x}{\alpha}$ follows from

$$\begin{aligned} &\left(\frac{\lambda}{r} (rb)^{\alpha+1} - (2\alpha + 1)(1 - rb) + (\alpha + 1) \left(\frac{\lambda}{r}\right)^{\frac{1}{\alpha+1}} (1 - rb)^{\frac{\alpha}{\alpha+1}} \right) \Big|_{rb=1-\frac{x}{\alpha}} \\ &= \frac{\lambda}{r} \left(1 - \frac{x}{\alpha}\right)^{\alpha+1} - \frac{2\alpha + 1}{\alpha} x + \frac{\alpha + 1}{\alpha} \left(\frac{\lambda \alpha}{r}\right)^{\frac{1}{\alpha+1}} x^{\frac{\alpha}{\alpha+1}} \\ &\xrightarrow{\alpha \rightarrow \infty} \frac{\lambda}{r} e^{-x} - x > 0. \end{aligned}$$

Second, we show that $\lim_{\alpha \rightarrow \infty} \alpha(1 - r\bar{d}) = \infty$. For any $M > 0$, $\lim_{\alpha \rightarrow \infty} \alpha(1 - r\bar{d}) > M$ because

$$(rd + \lambda r^\alpha d^{\alpha+1} - 1) \Big|_{rd=1-\frac{M}{\alpha}} = -\frac{M}{\alpha} + \lambda r^{-1} \left(1 - \frac{M}{\alpha}\right)^{\alpha+1} \xrightarrow{\alpha \rightarrow \infty} \lambda r^{-1} e^{-M} > 0.$$

Third, we show that $g(g(c)) < c$ for c slightly above $\bar{d}$. It is sufficient to show that $-g'(\bar{d}) < 1$, which follows from

$$-g'(\bar{d}) = \frac{(r + \lambda F(\bar{d}))^2}{\lambda f(\bar{d})} = \frac{1}{\alpha \lambda r^\alpha \bar{d}^{\alpha+1}} = \frac{1}{\alpha(1 - r\bar{d})} < 1,$$

where the inequality is shown in the second step.

Fourth, we show that $g(g(c)) < c$ for all $c \in (\bar{d}, b]$. Let $\check{d}$ be the smallest $c \in (\bar{d}, 1]$ such that $g(g(c)) = c$. Because $g'(g(\check{d}))g'(\check{d}) \geq 1$, and

$$\begin{aligned} g'(\check{d}) &= -\alpha^{-1} r^{-1} \lambda^{-\frac{1}{\alpha}} (1 - r\check{d})^{\frac{1}{\alpha}-1} \check{d}^{-\frac{1}{\alpha}-1}, \\ g'(g(\check{d})) &= -\frac{(r + \lambda F(g(g(\check{d}))))^2}{\lambda f(g(g(\check{d})))} = -\frac{(r + \lambda F(\check{d}))^2}{\lambda f(\check{d})} \\ &= -\left(r + \lambda r^\alpha \check{d}^\alpha\right)^2 \lambda^{-1} r^{-\alpha} \alpha^{-1} \check{d}^{1-\alpha}, \end{aligned}$$

we have

$$r \lambda^{1+\frac{1}{\alpha}} \alpha^2 (r\check{d})^\alpha (1 - r\check{d}) \leq (r + \lambda r^\alpha \check{d}^\alpha)^2 ((1 - r\check{d})/\check{d})^{\frac{1}{\alpha}} \leq (r + \lambda)^2.$$

It follows from $\lim_{\alpha \rightarrow \infty} \alpha(1 - r\bar{d}) = \infty$ that

$$\begin{aligned} \alpha^2 \left(1 - \frac{x}{\alpha}\right)^\alpha \left(\frac{x}{\alpha}\right) &\xrightarrow{\alpha \rightarrow \infty} \infty, \\ \lambda \alpha^2 (r\bar{d})^\alpha (1 - r\bar{d}) &= (\alpha(1 - r\bar{d}))^2 \bar{d}^{-1} \xrightarrow{\alpha \rightarrow \infty} \infty. \end{aligned}$$

Since $\alpha^2 (r\bar{d})^\alpha (1 - r\bar{d})$ remains bounded, $1 - \frac{x}{\alpha} < r\bar{d}$ for large α . This, together with $rb < 1 - \frac{x}{\alpha}$ shown in the first step, imply that $b < \check{d}$. Hence $g(g(c)) < c$ for all $c \in (\bar{d}, b]$. ■

Auxiliary Results for Appendix A

Lemma A.1 $dR(g(d)) \equiv d \int_0^{g(d)} (r^{-1} - c) f(c) dc$ is strictly concave in d .

Proof: The derivative of $R(g(d))$ is

$$\begin{aligned} \left(\int_0^{g(d)} (r^{-1} - c) f(c) dc \right)' &= -(r^{-1} - g(d)) f(g(d)) \frac{(r + \lambda F(g(d)))^2}{\lambda f(g(d))} \\ &= -(r^{-1} - g(d)) (r + \lambda F(g(d))) d^{-1} \lambda^{-1}. \end{aligned}$$

Hence the first derivative of $dR(g(d))$ is

$$\begin{aligned} &\int_0^{g(d)} (r^{-1} - c) f(c) dc - (r^{-1} - g(d)) (\lambda^{-1} r + F(g(d))) \\ &= \int_0^{g(d)} (r^{-1} - c) f(c) dc + (g(d) - r^{-1}) F(g(d)) + \lambda^{-1} r g(d) - \lambda^{-1} \\ &= \int_0^{g(d)} (g(d) - c) f(c) dc + \lambda^{-1} r g(d) - \lambda^{-1}. \end{aligned}$$

The second derivative of $dR(g(d))$ is

$$g'(d) (F(g(d)) + \lambda^{-1}r) < 0,$$

because $g'(d) < 0$. This verifies the strict concavity of $dR(g(d))$. ■

Lemma A.2 *If $v(\cdot)$ is decreasing and concave, then $(1 - rd)v(g(d))$ is concave in d under Assumption 1.*

Proof:

$$\begin{aligned} ((1 - rd)v(g(d)))' &= -rv(g(d)) + (1 - rd)v'(g(d))g'(d) \\ &= -rv(g(d)) - v'(g(d))(1 - rd)\frac{(r + \lambda F(g(d)))^2}{\lambda f(g(d))} \\ &= -rv(g(d)) - v'(g(d))\frac{F(g(d))(r + \lambda F(g(d)))}{f(g(d))}. \end{aligned}$$

Assumption 1 and $g'(d) < 0$ imply that $\frac{F(g(d))(r + \lambda F(g(d)))}{f(g(d))}$ decreases in d . Because both $-v(g(d))$ and $-v'(g(d))$ decrease in d , we know that $((1 - rd)v(g(d)))'$ decreases in d . This verifies concavity. ■

Lemma A.3 (Relaxed Problem) *When $d_0 \geq 0$ is not imposed, the solution to (3) is*

$$V_{\mathbb{R}}(d) = r^{-1}\lambda R(h(1 - rd)), \quad (9)$$

which is concave and strictly decreasing in d . The policy rule is

$$d_0 = d - h(1 - rd), \quad d_1 = h(1 - rd).$$

Proof: To show the monotonicity and concavity of $V_{\mathbb{R}}(\cdot)$, it is equivalent to show that $V'_{\mathbb{R}}(d)$ is negative and decreasing in d .

$$V'_{\mathbb{R}}(d) = -\lambda R'(d_1)h'(1 - rd) = \frac{-R'(d_1)}{(F(d_1)d_1)'} = \frac{-(r^{-1} - d_1)}{d_1 + F(d_1)/f(d_1)},$$

which is negative and increasing in d_1 under Assumption 1. As d_1 decreases in d , $V'_{\mathbb{R}}(d)$ is decreasing in d .

Next we verify the Bellman equation (3). Pick a feasible $(\tilde{d}_0, \tilde{d}_1)$, and let $\tilde{d}_2 \equiv h(1 - r(\tilde{d}_1 + \tilde{d}_0))$. Substituting $r\tilde{d}_0 = 1 - r\tilde{d}_1 - \lambda\tilde{d}_2F(\tilde{d}_2)$ into the PK constraint yields

$$\frac{r}{r + \lambda F(\tilde{d}_1)}\lambda F(\tilde{d}_1)\tilde{d}_1 + \frac{\lambda F(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)}\lambda F(\tilde{d}_2)\tilde{d}_2 \leq 1 - rd. \quad (10)$$

The objective on the right side of (3) is

$$\begin{aligned} \frac{\lambda R(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)} + \frac{\lambda F(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)} V_{\mathbb{R}}(\tilde{d}_1 + \tilde{d}_0) &= \frac{\lambda R(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)} + \frac{\lambda F(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)} r^{-1} \lambda R(\tilde{d}_2) \\ &= \frac{\lambda R(h(\tilde{x}_1))}{r + \lambda F(\tilde{d}_1)} + \frac{\lambda F(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)} r^{-1} \lambda R(h(\tilde{x}_2)), \end{aligned}$$

where $\tilde{x}_i = \lambda \tilde{d}_i F(\tilde{d}_i)$, $i = 1, 2$. Because $R(h(\cdot))$ is concave,

$$\begin{aligned} \frac{\lambda R(h(\tilde{x}_1))}{r + \lambda F(\tilde{d}_1)} + \frac{\lambda F(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)} r^{-1} \lambda R(h(\tilde{x}_2)) &\leq r^{-1} \lambda R \left(h \left(\frac{r}{r + \lambda F(\tilde{d}_1)} \tilde{x}_1 + \frac{\lambda F(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)} \tilde{x}_2 \right) \right) \\ &\leq r^{-1} \lambda R(h(1 - rd)) = V_{\mathbb{R}}(d), \end{aligned}$$

where the second inequality follows from (10). This verifies the Bellman equation. $\blacksquare$

Lemma A.4 $\Gamma'(d)|_{d=\bar{d}} = -r^{-1} \lambda F(\bar{d}) < 0$, where $\Gamma(d) \equiv d \lambda R(g(d)) + (1 - rd) r^{-1} \lambda R(\bar{d})$.

Proof: The proof of Lemma A.1 shows that $(R(g(d)))' = -(r^{-1} - g(d)) (r + \lambda F(g(d))) d^{-1} \lambda^{-1}$, hence

$$\begin{aligned} \Gamma'(d)|_{d=\bar{d}} &= -(r^{-1} - g(\bar{d})) (r + \lambda F(g(\bar{d}))) + \lambda R(g(\bar{d})) - \lambda R(\bar{d}) \\ &= -1 + \bar{d}r - r^{-1} \lambda F(\bar{d}) + \lambda \bar{d} F(\bar{d}) = -r^{-1} \lambda F(\bar{d}) < 0. \end{aligned}$$

$\blacksquare$

Lemma A.5 $\Omega'(d)|_{d=\bar{d}} = \left(1 - \frac{F(\bar{d})}{\bar{d}f(\bar{d})}\right) (\Gamma'(d)|_{d=\bar{d}})$, where

$$\Omega(d) \equiv d \lambda R(g(d)) + (1 - rd) \Gamma(g(d)),$$

and $\Gamma(\cdot)$ is defined in Lemma A.4. Hence $\Omega'(d)|_{d=\bar{d}} \geq 0$ if and only if $F(\bar{d}) \geq \bar{d}f(\bar{d})$.

Proof: $\Omega'(d)|_{d=\bar{d}} = (1 + (1 - r\bar{d})g'(\bar{d})) (\Gamma'(d)|_{d=\bar{d}})$. It follows from

$$1 - r\bar{d} = \frac{\lambda F(\bar{d})}{r + \lambda F(\bar{d})}, \quad g'(\bar{d}) = -\frac{(r + \lambda F(\bar{d}))^2}{\lambda f(\bar{d})},$$

that

$$1 + (1 - r\bar{d})g'(\bar{d}) = 1 - \frac{F(\bar{d}) (r + \lambda F(\bar{d}))}{f(\bar{d})} = 1 - \frac{F(\bar{d})}{\bar{d}f(\bar{d})}.$$

$\blacksquare$

Appendix B: Extensions

Allowing Value to Vary with Cost

Rather than taking the (per instant) social value of each innovation to be 1, suppose that value is related to cost according to $c = H(v)$. The quality of the product increases by v every time an innovation is made. To simplify the analysis, we assume that a better idea (i.e., that with a higher v) always implies a higher rate of return $\frac{v}{c}$.

Assumption 2 $\frac{v}{H(v)}$ increases in v .

Consider a truth-telling contract where the planner asks the innovator to report his type v . The planner chooses a subset $A \subseteq \mathbb{R}_+$ and recommends investment if and only if $v \in A$. If the innovator's reported v belongs to A , then $d_1(v)$ is the duration offered, $\Lambda(v)$ is the expected profits from contributions by other innovators, and $\Phi(v)$ is the expected fees charged to the innovator. If v does not belong to A , then there is no reward or punishment (see Footnote 6 for an explanation).

First, as in Scotchmer (1999), Assumption 2 implies that $A = [\bar{v}, \infty)$ for some $\bar{v} \geq 0$. Equivalently, we show that if $v_1 \in A$ and $v_1 < v_2$, then $v_2 \in A$. By contradiction, suppose $v_2 \notin A$. Because v_1 prefers to invest, $d_1(v_1)v_1 \geq H(v_1)$. Therefore a contradiction arises from

$$\begin{aligned} 0 &\leq v_1 \left(d_1(v_1) - \frac{H(v_1)}{v_1} \right) + \Lambda(v_1) - \Phi(v_1) \leq v_2 \left(d_1(v_1) - \frac{H(v_1)}{v_1} \right) + \Lambda(v_1) - \Phi(v_1) \\ &< v_2 \left(d_1(v_1) - \frac{H(v_2)}{v_2} \right) + \Lambda(v_1) - \Phi(v_1) \leq 0, \end{aligned}$$

where the third inequality follows from Assumption 2. The last inequality states that the payoff of a type v_2 innovator must be nonpositive if he reports v_1 , since he is recommended to not invest and receive nothing. Choosing $\bar{v}$ as the minimum in A finishes the proof.

Second, the optimal contract satisfies $d_1(v) = \frac{H(\bar{v})}{\bar{v}}$ for all $v \geq \bar{v}$. Suppose an incentive compatible contract recommends investment in $[\bar{v}, \infty)$. That type $\bar{v}$ invests requires $d_1(\bar{v}) \geq \frac{H(\bar{v})}{\bar{v}}$. Moreover, incentive compatibility implies that the duration $d_1(v)$ is increasing in v . Hence $d_1(v) \geq \frac{H(\bar{v})}{\bar{v}}$ for all $v \geq \bar{v}$ in any incentive compatible contract. As in Scotchmer (1999), we argue that the contract $\{d_1(v) = \frac{H(\bar{v})}{\bar{v}}, \Lambda(v) = \Phi(v), \forall v \geq \bar{v}\}$ is optimal. It is trivially incentive compatible since the duration promise is independent of report. More importantly, this contract minimizes the use of limited market time.

Third, we show how offering a duration $d_1 = \frac{H(\bar{v})}{\bar{v}}$ in this environment can be mapped into our problem with fixed quality improvement. Let $\theta = \frac{H(v)}{v}$ and define $v(\theta)$ to be its inverse function. Because θ is monotonically decreasing in v , the inverse function $v(\cdot)$ is well defined, and $v \geq \bar{v}$ if and only if $\theta \leq d_1$. If we denote the density of θ as $f(\theta)$, then the planner's payoff from offering d_1 is

$$R(d_1) = \int_0^{d_1} (v(\theta) - H(v(\theta)))f(\theta)d\theta = \int_0^{d_1} (1 - \theta) v(\theta)f(\theta)d\theta.$$

Choosing d_1 to implement types $\theta \in [0, d_1]$ is formally equivalent to our problem to implement c , so long as we interpret $v(\theta)f(\theta)$ as the transformed density for θ and $v(\theta)f(\theta)$ has a monotonic reverse hazard rate. When the planner increases duration, she trades off developing innovations with less-and-less net social benefit against foreclosing future innovations.

Sharing may Arise with Endogenous Innovation Size

As in Hopenhayn et al. (2006), time is discrete (i.e., $t = 0, 1, \dots$) and there is one innovator each period. To achieve a quality improvement of size e , an innovator of type θ incurs cost $\frac{e^2}{2\theta}$. Type θ measures efficiency as higher θ implies lower marginal cost. If an innovator is promised duration D , then he chooses innovation size e to

$$\max_e De - \frac{e^2}{2\theta},$$

which yields $e = \theta D$. The social planner's gain from each innovation is $\sum_{t=0}^{\infty} e(1+r)^{-t} = \frac{(1+r)e}{r}$, where r is the discount rate.

To make our point in the simplest setting, we focus on the sharing between the first two innovators (i.e., innovators 0 and 1) in period $t = 1$. To do so, suppose the two innovators receive total duration $\bar{d}_0$ from period 2 onwards and $\bar{d}_0$ is fixed. In this setting, the allocation is described by a pair of duration promises for the two innovators starting from period 1, (d_0, d_1) , subject to $d_0 + d_1 = 1 + \frac{\bar{d}_0}{1+r}$. Given (d_0, d_1) , the quality improvements for the two innovators are $e_0 = \theta_0(1 + \frac{d_0}{1+r})$ and $e_1 = \theta_1 d_1$, respectively. The planner's optimization problem is

$$\begin{aligned} V(d) = \max_{d_0, d_1, e_0, e_1} & \quad \frac{(1+r)e_0}{r} - \frac{e_0^2}{2\theta_0} + \frac{1}{1+r} \left(\frac{(1+r)e_1}{r} - \frac{e_1^2}{2\theta_1} \right) \\ \text{subject to} & \quad e_0 = \theta_0 \left(1 + \frac{d_0}{1+r} \right), \quad e_1 = \theta_1 d_1, \quad d_0 + d_1 = 1 + \frac{\bar{d}_0}{1+r}. \end{aligned}$$

In this problem, the optimal sharing rule depends only on $\frac{\theta_1}{\theta_0}$ (see Figure 8):

- (i) when $\frac{\theta_1}{\theta_0}$ is sufficiently small, there is full exclusion of innovator 1 (i.e., $d_1 = 0$);
- (ii) when $\frac{\theta_1}{\theta_0}$ is sufficiently large, innovator 0 exits when 1 arrives (i.e., $d_0 = 0$);
- (iii) when $\frac{\theta_1}{\theta_0}$ is neither small nor large, there is sharing (i.e., $d_0 > 0$ and $d_1 > 0$).

If one innovator is extremely efficient and the other extremely inefficient, then it is optimal for the planner to assign rights solely to the efficient type. This is what happened in Hopenhayn et al. (2006), where large heterogeneity among innovators allows the planner to focus on policies with no sharing. In our paper, however, the heterogeneity is not as extreme and sharing is optimal.

Figure 8: Policy functions with endogenous improvement size.

Cutoff Rule

Suppose the planner chooses a subset $A \subseteq [0, r^{-1}]$ and recommends investment if and only if $c \in A$. If $c \in A$, then the innovator claims rights and pays fees. If $c \notin A$, then there is no reward or punishment (see Footnote 6 for an explanation).

We show that $A = [0, \bar{c}]$ for some $\bar{c}$. Let $\bar{c}$ be the maximum in A , and we show that $c \in A$ if $c < \bar{c}$. The discounted payoff of the innovator with cost $\bar{c}$ must be nonnegative for him to be willing to invest. The discounted payoff of an innovator with cost c is at least $\bar{c} - c > 0$ because he can always report $\bar{c}$. Hence the planner cannot recommend an innovator with cost c to not invest and receive nothing. That is, $c \in A$.

Type Dependencies and Welfare

We show that the planner cannot improve welfare by designing type-dependent policies that differentiate among the cost types who developed their ideas. Suppose the planner recommends investment if and only if $c \leq \bar{c}$, and conditional on c , the incumbents and the new innovator receive durations $d_0(c)$ and $d_1(c)$, respectively. Compare this type-dependent policy with a type-independent policy in which the incumbents and the new innovator always receive $d_0 = \frac{\int_0^{\bar{c}} d_0(c) f(c) dc}{F(\bar{c})}$ and $d_1 = \bar{c}$.

Lemma 2 *The type-independent policy achieves a weakly higher social welfare.*

Proof: First, we show that $d_1(c) \geq \bar{c}$, $\forall c \leq \bar{c}$.

$$\bar{c} - c \leq d_1(c) - c + E \int_t^\infty e^{-r(s-t)} I_s(i) (n_s - 1 - \phi_s) ds \leq d_1(c) - c,$$

where the second inequality follows from $E \int_t^\infty e^{-r(s-t)} I_s(i)(n_s - 1 - \phi_s) ds \leq 0$ (innovator i cannot get a positive profit by reporting c and not investing). Thus, $d_1(c) \geq \bar{c}$, $\forall c \leq \bar{c}$.

Second, the planner's expected continuation value after the innovation is $\int_0^{\bar{c}} V(d_0(c) + d_1(c))f(c)dc$, which is higher under (d_0, d_1) because the concavity of $V(\cdot)$ implies that

$$\int_0^{\bar{c}} V(d_0(c) + d_1(c))f(c)dc \leq F(\bar{c})V\left(\int_0^{\bar{c}} (d_0(c) + d_1(c))\frac{f(c)}{F(\bar{c})}dc\right) \leq F(\bar{c})V(d_0 + \bar{c}),$$

where the second inequality relies on $d_1(c) \geq \bar{c}$ and the monotonicity of $V(\cdot)$. $\blacksquare$

Optimality of $\hat{d} = d$ and $\dot{d} = 0$

If the planner can also adjust the incumbents' promise when no idea arrives or when an innovator arrives but does not invest, then the promise-keeping constraint is

$$rd = y + \lambda F(d_1)(d_0 - d) + \lambda(1 - F(d_1))(\hat{d} - d) + \dot{d}. \quad (11)$$

To understand this constraint, it would be illuminating to first consider the case of $\lambda = 0$, that is, the case with no arrival of new innovators. Rewrite (11) as $\dot{d} = rd - y$. If the incumbents are assigned no rights (i.e., $y = 0$), then their duration promise grows at the discount rate r . Otherwise if the incumbents are assigned rights $y > 0$, then $\dot{d}$ is deducted by y to break even. Similarly, if $\lambda > 0$, then $F(d_1)(d_0 - d) + (1 - F(d_1))(\hat{d} - d)$ is the incumbents' expected gain/loss upon the arrival of a new innovator; the social planner must deduct this gain/loss from $\dot{d}$ so that the contingent rights offered to incumbents equal d in expectation. That is

$$\dot{d} = rd - y - \left(\lambda F(d_1)(d_0 - d) + \lambda(1 - F(d_1))(\hat{d} - d) \right).$$

The dynamic programming problem is

$$\begin{aligned} rV(d) = \max_{d_0, d_1, \hat{d}, y, \dot{d}} & \quad \lambda R(d_1) + \lambda F(d_1)(V(d_1 + d_0) - V(d)) \\ & \quad + \lambda(1 - F(d_1))(V(\hat{d}) - V(d)) + V'(d)\dot{d}, \\ \text{subject to} & \quad (11). \end{aligned}$$

We have shown that $V(\cdot)$ is concave in the proof of Propositions 1 and 2. We can use the first-order conditions and the envelope condition to immediately conclude that

Lemma 3 *If $V(\cdot)$ is concave, then $\hat{d} = d$ and $\dot{d} = 0$.*

Proof: Let $\mu(d)$ be the Lagrange multiplier on the PK constraint (11). The first-order conditions for $\hat{d}$ and $\dot{d}$ are, respectively,

$$\lambda(1 - F(d_1))V'(\hat{d}) + \lambda(1 - F(d_1))\mu(d) = 0, \quad V'(d) + \mu(d) = 0,$$

which imply that $V'(\hat{d}) = V'(d)$, and hence $\hat{d} = d$. The envelope condition is

$$-(r + \lambda)V'(d) + V''(d)\dot{d} - (r + \lambda)\mu(d) = 0,$$

which implies that $\dot{d} = 0$ because $V'(d) + \mu(d) = 0$. $\blacksquare$

References

- Akcigit, U., Kerr, W. R., 2010. Growth through heterogeneous innovations. Harvard Business School Entrepreneurial Management Working Paper No. 11-044.
- Aoki, R., Hu, J.-L., 1999. Licensing vs. litigation: the effect of the legal system on incentives to innovate. *Journal of Economics and Management Strategy* 8, 133–160.
- Bebchuk, L. A., 1984. Litigation and settlement under imperfect information. *The RAND Journal of Economics* 15, 404–415.
- Bessen, J., 2004. Holdup and licensing of cumulative innovations with private information. *Economics Letters* 82, 321–326.
- Bhattacharyya, S., Lafontaine, F., 1995. Double-sided moral hazard and the nature of share contracts. *The RAND Journal of Economics* 26, 761–781.
- Chari, V., Golosov, M., Tsyvinski, A., 2011. Prizes and patents: Using market signals to provide incentives for innovations. *Journal of Economic Theory* forthcoming.
- Choi, J. P., 1998. Patent litigation as an information-transmission mechanism. *American Economic Review* 88, 1249–1263.
- Chou, T., Haller, H., 2007. The division of profit in sequential innovation for probabilistic patents. *Review of Law & Economics* 3.
- Gallini, N., 2012. Promoting competition by coordinating prices: When rivals share intellectual property. Working paper, University of British Columbia.
- Green, J. R., Scotchmer, S., 1995. On the division of profit in sequential innovation. *The RAND Journal of Economics* 26, 20–33.
- Henry, E., 2010. Promising the right prize. Working paper, London School of Business.
- Holmstrom, B., 1982. Moral hazard in teams. *Bell Journal of Economics, The RAND Corporation* 13 (2), 324–340.
- Hopenhayn, H., Llobet, G., Mitchell, M., 2006. Rewarding sequential innovators: Patents, prizes, and buyouts. *Journal of Political Economy* 115 (6), 1041–1068.
- Kremer, M., 1998. Patent buyouts: A mechanism for encouraging innovation. *Quarterly Journal of Economics* 113 (4), 1137–1167.
- Kremer, M., 2000. Creating markets for new vaccines part ii: Design issues. Tech. Rep. 7717, National Bureau of Economic Research, Inc.
- Lemley, M., Shapiro, C., 2005. Probabilistic patents. *Journal of Economic Perspectives* 19 (2), 75–98.

- Lerner, J., Stojwas, M., Tirole, J., 2007. The design of patent pools: The determinants of licensing rules. *The RAND Journal of Economics* 38 (3), 610–625.
- Lerner, J., Tirole, J., 2004. Efficient patent pools. *American Economic Review* 94 (3), 691–711.
- Lerner, J., Tirole, J., April 2008. Public policy toward patent pools. In: Adam B. Jaffe, J. L., Stern, S. (Eds.), *Innovation Policy and the Economy*. Vol. 8. University of Chicago Press, pp. 157–186.
- Meurer, M. J., 1989. The settlement of patent litigation. *The RAND Journal of Economics* 20, 77–91.
- Mill, J. S., 1883. *Principles of political economy: with some of their applications to social philosophy*. D. Appleton and Co.
- Nalebuff, B., 1987. Credible pretrial negotiation. *The RAND Journal of Economics* 18, 1198–1210.
- O'Donoghue, T., 1998. A patentability requirement for sequential innovation. *The RAND Journal of Economics* 29 (4), 654–679.
- O'Donoghue, T., Scotchmer, S., Thisse, J.-F., 1998. Patent breadth, patent life, and the pace of technological progress. *Journal of Economics & Management Strategy* 7 (1), 1–32.
- Posner, E. A., 2003. Economic analysis of contract law after three decades: Success or failure. *The Yale Law Journal* 112 (4), 829–880.
- Robinson, W. T., Kalyanaram, G., Urban, G., 1994. First-mover advantages from pioneering new markets: A survey of empirical evidence. *Review of Industrial Organization* 9 (1), 1–23.
- Scotchmer, S., 1996. Protecting early innovators: Should second-generation products be patentable? *The RAND Journal of Economics* 27 (2), 322–331.
- Scotchmer, S., 1999. On the optimality of the patent renewal system. *The RAND Journal of Economics* 30 (2), 181–196.
- Shapiro, C., 2003. Antitrust limits to patent settlements. *The RAND Journal of Economics* 34 (2), 391–411.
- Weyl, E. G., Tirole, J., 2011. Market power screens willingness-to-pay. SSRN eLibrary.