

Shared rights and technological progress*

Matthew Mitchell[†] and Yuzhe Zhang[‡]

January 30, 2014

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. Because such opportunities are limited, awarding rights to one innovator reduces the rights awarded 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 are grateful to the editor Guido Menzio and two anonymous referees, whose comments on a previous version of this paper helped us revise the paper substantially. We are also grateful to B. Ravikumar, Hugo Hopenhayn, Qi Li 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, 105 St. George St., Toronto, ON M5S 3E6, Canada. Email: matthew.mitchell@rotman.utoronto.ca. Tel.: 416-946-3149.

[‡]Department of Economics, Texas A&M University, College Station, TX, 77843. Email: yuzhe-zhang@econmail.tamu.edu. Tel.: 319-321-1897.

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.

In the information technology sector, for instance, the DVD player relies on a pool of patents owned by many rights holders, who share the licensing fees that are paid to the pool. Wireless networking standards like Bluetooth and the standard 802.11 for

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.”

wireless routers are made up of many patents held by disparate innovators, who share royalties from the standard. Other examples include the MPEG and MP3 standards for file compression, and the 4G standard for cellular communication. Our model, then, provides a rationale for these observed practices. Moreover, it addresses questions related to the welfare loss if sharing were either difficult or ruled out by competition policy.²

Our model features a sequence of innovators who make quality improvements on one another, i.e., later innovators stand on the shoulders of those who came before. In the wireless networking example, for instance, standards undergo frequent revisions to incorporate innovations that improve their functionality. Ideas for making improvements arrive randomly. An innovator 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 matches the notion that the technologies we mention involve many innovations, and determining the marginal contribution of any given innovation is not possible. 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

²While patent pools are currently allowed in the U.S., they have not always been, on the grounds that they might be an anticompetitive way for rights holders to increase their market power. Our model is silent on this possible cost of sharing, but does illuminate a key benefit of sharing.

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 pool must be ever growing since our sharing rule involves an ever-improving product. Firms do not simply sell their rights but rather share in the profits of future innovations. 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.

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 equity share 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 implementation of shared rights in our model is through a lottery among the risk-neutral innovators who share the rights. A winner keeps all future profits promised to both winner and loser. This implementation makes it clear that sharing has the benefit of being 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. If sharing is possible, improvements arrive at a constant rate and different arrivals generate equal expected net social benefits.

The idea of intellectual property rights as random variables matches the notion of weak patent rights in the literature, where the randomization is commonly interpreted as litigation. This interpretation shows the sense in which shared rights, while possibly having benefits, may come with substantial costs. Meurer and Bessen (2005) report that costs for these suits range from nearly \$500,000 for cases with \$1 million at risk, up to nearly \$4 million for cases with more than \$25 million at risk. One could also imagine that the licensing contract that implements the optimal sharing allocation might be impossible to negotiate in many environments. Rather than modeling the contracting imperfections 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. The allocation without sharing can be thought of as a sequence of patents that do not infringe on prior art, so that no issue of licensing arises. When sharing is avoided, the allocation of rights is distorted. The arrival of improvements follows a less smooth path, bypassing some higher net benefit ideas at the beginning and implementing other less attractive ideas later on. We show that, in such cases, the lack of sharing can lead to perpetual cycles in technological progress, 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. Our model wedds the sequential innovation structure of Hopenhayn et al. (2006) and O’Donoghue et al. (1998) with an underlying idea process in keeping with Scotchmer (1999). In contrast with Hopenhayn et al. (2006), where the sequence of innovations has types that are associated with lower total and marginal cost, our model has a type structure similar to Scotchmer (1999), where the size and cost of the innovation are both “hard wired” to the innovator’s type. This means that, whereas Hopenhayn et al. (2006) assumes that more protection improves a given idea’s development at the intensive margin, and in a sufficiently strong way to make rights exclusive, in our model greater rights granted by the planner can only increase the extensive margin of innovation. We show that this difference leads to an optimal allocations featuring sharing, patentability, and infringement from a problem with explicit incomplete information, and therefore provides an informational foundation for the optimal policies studied in many other papers on cumulative innovation that is absent in Hopenhayn et al. (2006).

Our model also provides a new interpretation of patent pools and probabilistic patents. Existing models of pools, such as Lerner and Tirole (2004) and Gallini (2012) treat a pool as similar to mergers, focusing on the conduct of the pool vis-a-vis its customers. Our interpretation offers a different view of patent pools: patent pools split rights among competing innovators in a way that gives incentives for new members to produce innovations that enhance the pool, which connects the sharing rules for patent pools with the incentive-to-innovate issue at the heart of the optimal patent literature. The 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, 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. In this sense our paper provides an underpinning for models that assume probabilistic patent rights.

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 Technology and preferences

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 innovator who receives the idea privately knows its arrival: the innovator's identity becomes public only if he reports the idea to the planner.

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 $n - p$. If the consumer does not consume the good, his utility is zero. Firms have no cost of production. If a firm has the exclusive rights to sell one unit of the good (i.e., the firm is a monopolist), it sets price p to solve the following profit maximization problem

$$\begin{aligned} \max_p \quad & p \\ \text{subject to} \quad & n - p \geq 0. \end{aligned}$$

The firm makes n units of profits by charging the optimal price $p = n$. 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.

2.2 Information and public history

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.³ We therefore focus on policies that reward innovators with the opportunity to market their innovations, which mimic the observed practice of patent policies.

The moral hazard assumption is natural for cases like the information technology examples that motivate our study of sharing. In those cases it is impractical to assess the contribution of any given patent to the pool; doing so would require selling a variety of versions of the product side by side to infer the marginal contribution of each of the many innovations in the pool.⁴ Indeed, the determination of the shares of pool royalties is very contentious in practice, underscoring the difficulty in determining the value of each contribution.

In our model, an innovator 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 innovator to report the quality, to cross-check past innovators' contributions. However, we assume that, although the quality and market profits are observable by the innovator, 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 an innovator will not face punishment by the planner (or legal action by other innovators) 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

³Without 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.

⁴The MPEG pool has more than 1000 patents, for instance. As a result, one would need to offer many different versions, and observe cash flows very frequently, to back out the contribution of individual innovators.

⁵Kremer (2000), Chari et al. (2012), Weyl and Tirole (2012) and Henry (2010) study situations where the planner uses market signals.

following Kremer (1998).

The history at every point in time is defined as follows. Let m_t be the number of innovators who have reported an idea up to time t . For the i th innovator, $i \in \{1, 2, \dots, m_t\}$, denote his (reported) arrival time as t_i and his (reported) development cost as c_i . The *public history* of reports at t is

$$h_t \equiv \{t, m_t, (t_i, c_i)_{i=1}^{m_t}\},$$

which records the number of innovators who have reported ideas, their arrival times, and cost types.⁶ Because we have ruled out cross-reporting, history does not contain any information on the quality of the product or market profits. For later convenience, let $h_{t-} \equiv \lim_{s \uparrow t} h_s$ be the history immediately before t . That is, h_{t-} is the history at the beginning of t and before the realization of any time- t shocks; h_t equals h_{t-} plus the arrival (if any) reported in time t . Let H_t be the set of all histories at time t , i.e.,

$$H_t \equiv \{h_t : t \geq 0, m_t \text{ is an integer}, 0 \leq t_1 < t_2 < \dots < t_{m_t} \leq t, (c_i)_{i=1}^{m_t} \geq 0\}.$$

Next we describe the evolution of h_t . If $\epsilon > 0$ is a small positive number, then the number of new arrivals at the beginning of $t + \epsilon$ is either zero or one. With probability $1 - \lambda\epsilon$, no idea arrives and $m_{t+\epsilon} = m_t$. In this case the new history is $h_{t+\epsilon} = \{t + \epsilon, m_t, (t_i, c_i)_{i=1}^{m_t}\}$. With probability $\lambda\epsilon$, one idea arrives and $m_{t+\epsilon} = m_t + 1$. If the new innovator reports cost c , then the new history is $h_{t+\epsilon} = \{t + \epsilon, m_t + 1, ((t_i, c_i)_{i=1}^{m_t}, (t + \epsilon, c))\}$.

2.3 Allocations

The timing within date t is as follows. First, the uncertainty about whether a new idea will arrive is realized. If a new idea arrives, then the innovator with the idea draws a private cost c . If the innovator reports cost \tilde{c} to the planner, a new history h_t is generated by $h_t = \{h_{t-}, (t, \tilde{c})\}$. The innovator decides whether to report truthfully after observing the public history h_{t-} and his private cost. If no idea arrives, $h_t = h_{t-}$. Second, conditional on h_t , the planner makes a recommendation to the innovator whether to pay the investment cost. Denote this recommendation as $C(h_t) \in \{0, 1\}$, where 0 means

⁶The private history of an innovator who arrives at time t is h_t plus his privately observed cost.

“do not invest” and 1 means “invest.” The innovator decides whether to follow the recommendation. If he invests, the quality of the leading-edge product is increased by one; otherwise, the quality remains unchanged. Third, conditional on h_t , the planner determines the identity of an exclusive rights holder for the leading-edge product through an indicator function $I(h_t, \cdot) : \{1, 2, \dots, m_t\} \rightarrow \{0, 1\}$; i.e., innovator i holds the exclusive rights if and only if $I(h_t, i) = 1$. At most one of $\{I(h_t, i) : 1 \leq i \leq m_t\}$ can be 1 and the planner may allow no innovator to profit in a given instant by choosing $I(h_t, i) = 0$ for all i . The planner also charges a fee, $\phi(h_t) \in \mathbb{R}$, in exchange for the rights granted; fees collected are rebated to the consumer lump sum. Finally, production and consumption take place.

Although the planner does not observe the quality of the product, she can infer it if all innovators report truthfully and follow her recommendations. This inferred quality is $n_t \equiv \sum_{i=1}^{m_t} C(h_{t_i})$.

Definition 1 A **mechanism** (or an **allocation**) is a history-dependent plan $(C(h_t), I(h_t, i), \phi(h_t))$ for all $t \geq 0$, all $h_t \in H_t$, and all $i \in \{1, 2, \dots, m_t\}$.

At time zero, the planner designs the mechanism and commits to it. We study how the mechanism affects innovators’ incentives next.

2.4 Innovators’ strategies and payoffs

At time t there may be two types of innovators: an innovator with no idea, and an innovator i with an idea and a development cost of c . First, we consider innovators’ payoffs when they report truthfully and always follow the planner’s recommendation. If an innovator has no idea and reports “no idea,” then he receives no rights and pays no fees and therefore his payoff is zero. If innovator i reports his cost c truthfully, receives a recommendation $C(h_t) = 1$, and follows the recommendation to invest, then his payoff is

$$E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i)(n_s - \phi(h_s)) ds \middle| h_t \right] - c.$$

Here the expectation is taken with respect to uncertain future histories h_s conditional on the history h_t and under the assumption that future innovators would report truthfully and follow the planner's recommendations too. Recall that we have described the law of motion of $\{h_s; s \geq t\}$ earlier. If innovator i reports c truthfully, receives $C(h_t) = 0$, and follows the recommendation that he should not invest, then his payoff is

$$E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i)(n_s - \phi(h_s)) ds \middle| h_t \right].$$

Second, we consider innovator i 's payoff when he either misreports his type or does not follow the planner's recommendation, or both. If innovator i reports $\tilde{c} \neq c$, then the history at t becomes

$$\tilde{h}_t \equiv \{h_{t-}, (t, \tilde{c})\}.$$

Since c would never be revealed, the after- t history under misreporting follows the same law of motion as under truth telling, except for a different initial condition $\tilde{h}_t \neq h_t$. We denote the after- t history as $\{\tilde{h}_s; s \geq t\}$ and will elaborate on it further in the following. If $C(\tilde{h}_t) = 1$ and the innovator follows the recommendation to invest, then his payoff is

$$U(c, \tilde{c}, 1) \equiv E \left[\int_t^\infty e^{-r(s-t)} I(\tilde{h}_s, i)(\tilde{n}_s - \phi(\tilde{h}_s)) ds \middle| \tilde{h}_t \right] - c, \quad (1)$$

where the first two arguments of U are the innovator's true cost c and reported cost \tilde{c} , respectively, and the last argument 1 means that the innovator "invests." If $C(\tilde{h}_t) = 1$ but the innovator does not invest, then future qualities will all be lower by one, and therefore the payoff to shirking, which is independent of c , is

$$U(\tilde{c}, 0) \equiv E \left[\int_t^\infty e^{-r(s-t)} I(\tilde{h}_s, i)(\tilde{n}_s - 1 - \phi(\tilde{h}_s)) ds \middle| \tilde{h}_t \right]. \quad (2)$$

Note that, despite a lower quality, the after- t history $\{\tilde{h}_s; s \geq t\}$ under shirking in (2) is identical to that in (1). This is because future innovators are unaware of any previous shirking when they report their types and make investment decisions; they only detect the shirking after they are granted exclusive rights to sell the leading-edge product and observe cash flows. Therefore, innovator i 's shirking (and a lower quality thereafter) does not affect future innovators' strategies (i.e., future innovators still report their types truthfully and follow the planner's recommendations).

If $C(\tilde{h}_t) = 0$, the innovator's payoff when he does and does not invest are, respectively,

$$U(c, \tilde{c}, 1) \equiv E \left[\int_t^\infty e^{-r(s-t)} I(\tilde{h}_s, i)(\tilde{n}_s + 1 - \phi(\tilde{h}_s)) ds \middle| \tilde{h}_t \right] - c, \quad (3)$$

$$U(\tilde{c}, 0) \equiv E \left[\int_t^\infty e^{-r(s-t)} I(\tilde{h}_s, i)(\tilde{n}_s - \phi(\tilde{h}_s)) ds \middle| \tilde{h}_t \right]. \quad (4)$$

Again, the after- t histories $\{\tilde{h}_s; s \geq t\}$ are the same in (3) and (4).

Finally, if innovator i hides his idea (i.e., he has an idea but reports “no idea”), then his payoff is zero. On the other hand, if an innovator with no idea reports an idea with cost \tilde{c} , then his payoff is $U(\tilde{c}, 0)$ defined in either (2) or (4) since he cannot invest.

2.5 Incentive compatibility and the planner's problem

Definition 2 A mechanism is *incentive compatible* (IC), if for any public history h_{t-} and any private shock an innovator receives at time t , the innovator prefers to report his type truthfully and follow the planner's investment recommendation. In particular, if he has no idea, then

$$0 \geq U(\tilde{c}, 0), \quad \forall \tilde{c}. \quad (5)$$

If he has an idea with cost c and $C(h_t) = 1$ for $h_t = \{h_{t-}, (t, c)\}$, then

$$U(c, c, 1) \geq \max \{0, U(c, \tilde{c}, 1), U(\tilde{c}, 0)\}, \quad \forall \tilde{c}. \quad (6)$$

If $C(h_t) = 0$, then

$$U(c, 0) \geq \max \{0, U(c, \tilde{c}, 1), U(\tilde{c}, 0)\}, \quad \forall \tilde{c}. \quad (7)$$

Because the planner has commitment, the revelation principle applies and we could restrict attention to IC mechanisms. We turn to the planner's problem next.

The discounted 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 the price is zero). Therefore, the planner's expected payoff conditional on the arrival of an idea at h_{t-} is

$$\int_0^\infty C(h_t)(r^{-1} - c)f(c)dc, \quad \text{where } h_t = \{h_{t-}, (t, c)\}.$$

Note that the fees do not enter the social welfare function: they increase the consumer surplus but decrease firms' profits. The planner's total discounted payoff under a mechanism $\sigma \equiv \{(C(h_t), I(h_t, \cdot), \phi(h_t)); t \geq 0, h_t \in H_t\}$ is

$$V(\sigma) \equiv E \left[\sum_{i=1}^{\infty} e^{-rt_i} \left(\int_0^{\infty} C(h_{t_i})(r^{-1} - c) f(c) dc \right) \right].$$

The planner's problem is to design an IC mechanism σ to maximize $V(\sigma)$.

2.6 Simplification

This subsection presents a few properties of IC mechanisms, which we will use to simplify the planner's problem later on. Suppose the history is $h_t = \{h_{t-}, (t, c)\}$, i.e., innovator i reports an idea with cost c at time t .

Lemma 1 *If $C(h_t) = 0$, then $U(c, 0) = 0$.*

There can be neither reward nor punishment for an innovator who receives a recommendation that he should not invest. If there is a reward, an innovator with no idea would pretend to have an idea. If there is a punishment, an innovator with an idea would hide the arrival of his idea.

Lemma 2 *If $C(h_t) = 1$ and $c' < c$, then $C(h'_t) = 1$ for $h'_t = \{h_{t-}, (t, c')\}$.*

This result is not surprising: if the planner implements an idea c , then she should implement the idea c' that is even better (in the sense that it requires a lower cost). The recommendation for investment must be a cutoff rule. Define the cutoff

$$\bar{c}(h_{t-}) \equiv \sup\{c : C(h_t) = 1 \text{ for } h_t = \{h_{t-}, (t, c)\}\},$$

and innovator i is recommended to invest if and only if c is below $\bar{c}(h_{t-})$.

Lemma 3 *For any IC mechanism $\sigma \equiv \{C, I, \phi\}$ and any history h_t at which innovator i is recommended not to invest (i.e., $C(h_t) = 0$), a modified mechanism $\tilde{\sigma} \equiv \{C, \tilde{I}, \phi\}$ is IC with $V(\tilde{\sigma}) = V(\sigma)$, where*

$$\tilde{I}(h_s, j) = \begin{cases} 0, & \text{if } s \geq t, \text{ history } h_s \text{ follows } h_t, \text{ and } j = i; \\ I(h_s, j), & \text{otherwise.} \end{cases}$$

Lemma 3 allows us to restrict attention to mechanisms that grant no rights to innovators who receive recommendations that they should not invest. Although $V(\tilde{\sigma}) = V(\sigma)$, we show (after Lemma 5) that the modified $\tilde{\sigma}$ may allow for further modifications that are welfare improving.

If innovator i is recommended to invest, i.e., $C(h_t) = 1$, then define his expected discounted *duration* of rights as

$$d_1(h_t) \equiv E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i) ds \middle| h_t \right].$$

Lemma 4 *If $C(h_t) = 1$, then $d_1(h_t) \geq \bar{c}(h_{t-})$.*

The intuition for Lemma 4 can be easily obtained if $c = \bar{c}(h_{t-})$. In this case, the innovator with cost $\bar{c}(h_{t-})$ is recommended to invest. If he does so, his payoff is

$$U(c, c, 1) = E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i) (n_s - \phi(h_s)) ds \middle| h_t \right] - \bar{c}(h_{t-}).$$

If he reports c but does not invest, his payoff is

$$\begin{aligned} U(c, 0) &= E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i) (n_s - 1 - \phi(h_s)) ds \middle| h_t \right] \\ &= E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i) (n_s - \phi(h_s)) ds \middle| h_t \right] - d_1(h_t). \end{aligned}$$

Then $d_1(h_t) \geq \bar{c}(h_{t-})$ must hold for him to be willing to invest.

Lemma 4 shows the sense in which the exclusive rights to sell the leading-edge product are scarce. Recall that the planner's payoff conditional on an arrival is

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

The planner would like to choose $\bar{c} = r^{-1}$: every idea whose cost is less than its social benefit r^{-1} ought to be implemented. In this case, however, Lemma 4 requires that each innovator receive a discounted duration of at least r^{-1} . In other words, each innovator must be made a perpetual rights holder after his arrival. This certainly generates conflicts between early and late arrivals.

Lemma 5 *For any IC mechanism $\sigma \equiv \{C, I, \phi\}$ and any history h_{t-} at which innovator i faces an investment cutoff \bar{c} , there exists a modified IC mechanism $\tilde{\sigma} \equiv \{C, \tilde{I}, \tilde{\phi}\}$ with $V(\tilde{\sigma}) = V(\sigma)$ and*

(i) $\tilde{I}(h_s, j) = I(h_s, j)$, $\forall h_s, \forall j \neq i$;

(ii) $\tilde{\phi}(h_s) = \phi(h_s)$ and $\tilde{I}(h_s, i) = I(h_s, i)$ except on

$$\{h_s : h_s \text{ follows } h_t = \{h_{t-}, (t, c)\} \text{ for some } c \leq \bar{c} \text{ and } I(h_s, i) = 1\};$$

(iii) if $h_t = \{h_{t-}, (t, c)\}$ for some $c \leq \bar{c}$, then

$$E \left[\int_t^\infty e^{-r(s-t)} \tilde{I}(h_s, i) ds \mid h_t \right] = \bar{c}.$$

The proof of Lemma 5 shows how to modify the duration promises such that investing innovators receive exactly \bar{c} and modify the fees such that the mechanism remains IC.

Lemma 5 allows us to restrict attention to mechanisms that grant exactly \bar{c} to innovators who are recommended to invest. Recall that Lemma 3 allows us to restrict attention to mechanisms that grant no rights to innovators who receive recommendations that they should not invest. Although the planner's payoff is unaffected by the modifications in Lemmas 3 and 5 (i.e., $V(\tilde{\sigma}) = V(\sigma)$), the planner may have another modification in which she uses the rights saved from innovator i to reward later innovators. This last modification raises the cutoff for later innovators and is welfare improving.

Given the above lemmas, we simplify the mechanism to $\{\bar{c}(h_{t-}), I(h_t, \cdot); t \geq 0, h_t \in H_t\}$. We replace $C(h_t)$ by $\bar{c}(h_{t-})$ because the recommendation for investment is a cutoff rule. We eliminate $\phi(h_t)$ from the mechanism because the fees are welfare neutral and, as the proof of Lemma 5 shows, they can always be designed such that the mechanism is IC. To summarize, the planner's problem is

$$\begin{aligned} & \max_{\substack{\bar{c}(h_{t-}), I(h_t, \cdot) \\ t \geq 0, h_t \in H_t}} E \left[\sum_{i=1}^{\infty} e^{-rt_i} R(\bar{c}(h_{t_i-})) \right] \\ \text{subject to} \quad & \bar{c}(h_{t_i-}) = E \left[\int_{t_i}^{\infty} e^{-r(s-t_i)} I(h_s, i) ds \mid h_{t_i} \right], i = 1, 2, \dots, \end{aligned}$$

where $\bar{c}(h_{t_i-}) = E \left[\int_{t_i}^{\infty} e^{-r(s-t_i)} I(h_s, i) ds \mid h_{t_i} \right]$ is both the cutoff cost and the promised duration if innovator i claims a cost below $\bar{c}(h_{t_i-})$. All innovators with a cost below $\bar{c}(h_{t_i-})$ at t_i receive the same duration $\bar{c}(h_{t_i-})$. Note that we do not deal with IC constraints from now on, since they are always satisfied if the planner chooses the appropriate fees.

3 Solving the planner's problem

3.1 A recursive formulation

At history h_t , the future duration promised to innovator i , who arrives before t (i.e., $t_i \leq t$), is $E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i) ds \mid h_t \right]$. The sum of durations promised to all existing innovators is

$$d \equiv \sum_{i, t_i \leq t} E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i) ds \mid h_t \right],$$

while the duration available to subsequent innovators is $r^{-1} - d$. The planner's problem starting from time t onward is to efficiently allocate the remaining duration $r^{-1} - d$ among future innovators to maximize

$$V(h_t) \equiv E \left[\sum_{i, t_i > t}^{\infty} e^{-r(t_i-t)} R(\bar{c}(h_{t_i-})) \mid h_t \right].$$

We make two remarks about the maximized welfare $V(h_t)$. First, d matters for the above maximization problem. A greater d 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. Second, $V(h_t)$ relies on the history h_t only through d . The planner, whose objective is to maximize the welfare gains of future innovations, does not care about the ownership structure behind d (i.e., which past innovator owns how much of d). She only cares about $r^{-1} - d$, the total duration that she can allocate to future innovators. Formally, we have

Lemma 6 *If h_t and \tilde{h}_s are two histories such that $d(h_t) = d(\tilde{h}_s)$, then $V(h_t) = V(\tilde{h}_s)$.*

Put differently, d is a sufficient statistic for the history h_t . Summarizing histories in this way allows us to rewrite the planner's value function as $V(d)$.

Next, we transform the planner's sequence problem into a dynamic programming problem, using d as the state variable. We start with a description of the evolution of the state variable in this dynamic programming problem. Suppose the planner finds herself with an outstanding duration promise of d at time t . Let $\epsilon > 0$ be a small number. The history $h_{t+\epsilon}$ at time $t + \epsilon$ can be one of three cases: (1) no idea arrives; (2)

an idea with cost $c > \bar{c}(h_{t+\epsilon-})$ arrives; or (3) an idea with cost $c \leq \bar{c}(h_{t+\epsilon-})$ arrives. To simplify notation, we suppress the dependence of $h_{t+\epsilon}$ on h_t and denote the history $h_{t+\epsilon}$ as “no idea”, “do not invest”, and “invest”, respectively. In the first two cases, the set of rights holders does not change from t to $t + \epsilon$ (in the second case the innovator with cost $c > \bar{c}(h_{t+\epsilon-})$ does not receive rights), while in the third case, the new innovator with cost $c \leq \bar{c}(h_{t+\epsilon-})$ becomes a new rights holder. The state variables at histories “no idea” and “do not invest” are, respectively,

$$\begin{aligned} d(\text{no idea}) &\equiv \sum_{i, t_i \leq t} E \left[\int_{t+\epsilon}^{\infty} e^{-r(s-(t+\epsilon))} I(h_s, i) ds \mid \text{no idea} \right], \\ d(c) &\equiv \sum_{i, t_i \leq t} E \left[\int_{t+\epsilon}^{\infty} e^{-r(s-(t+\epsilon))} I(h_s, i) ds \mid \text{do not invest} \right], \quad \forall c > \bar{c}(h_{t+\epsilon-}). \end{aligned}$$

Again, we have suppressed the dependence of $d(\text{no idea})$ and $d(c)$ on h_t . Because there is one more rights holder at the history “invest,” the state variable is the sum of the duration d_0 for innovators who arrive before t and the duration d_1 for the new innovator $m_t + 1$ who arrives at $t + \epsilon$. That is,

$$\begin{aligned} d_0(c) &\equiv \sum_{i, t_i \leq t} E \left[\int_{t+\epsilon}^{\infty} e^{-r(s-(t+\epsilon))} I(h_s, i) ds \mid \text{invest} \right], \quad \forall c \leq \bar{c}(h_{t+\epsilon-}), \\ d_1 &\equiv E \left[\int_{t+\epsilon}^{\infty} e^{-r(s-(t+\epsilon))} I(h_s, m_t + 1) ds \mid \text{invest} \right]. \end{aligned}$$

Recall from subsection 2.6 that $d_1 = \bar{c}(h_{t+\epsilon-})$ does not depend on c .

Suppose the planner offers to the incumbents a fraction $y \in [0, 1]$ of the instants in $[t, t + \epsilon]$. 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. The following promise-keeping (PK) constraint ensures that the planner delivers the promised d to the incumbents:

$$d = y\epsilon + e^{-r\epsilon} \left[(1 - \lambda\epsilon)d(\text{no idea}) + \lambda\epsilon \left[\int_0^{d_1} d_0(c)f(c)dc + \int_{d_1}^{\infty} d(c)f(c)dc \right] \right], \quad (8)$$

where $\lambda\epsilon$ is the probability of an arrival, and conditional on the arrival $\int_0^{d_1} d_0(c)f(c)dc$ and $\int_{d_1}^{\infty} d(c)f(c)dc$ are the duration for the incumbents when the new innovator does

and does not invest, respectively. The Bellman equation for the value function is

$$V(d) = \max e^{-r\epsilon} \left[(1 - \lambda\epsilon)V(d(\text{no idea})) + \lambda\epsilon \left[\int_{d_1}^{\infty} V(d(c))f(c)dc \right. \right. \\ \left. \left. + \int_0^{d_1} (r^{-1} - c + V(d_0(c) + d_1)) f(c)dc \right] \right]. \quad (9)$$

Because r^{-1} already accounts for future contributions of the innovation developed at $t + \epsilon$, the continuation value $V(d_0(c) + d_1)$ must exclude those contributions to avoid double counting. This is consistent with the definition of $V(\cdot)$ as the discounted value of future, but not existing, innovations. As in standard dynamic contracting problems, the planner maximizes the right-hand side of (9) by choosing y and the future state variables (i.e., promises $d(\text{no idea})$, $d(c)$, $d_0(c)$, and d_1) subject to the PK constraint.

3.2 A restricted problem

Problem (9) is not easily solvable due to its large number of control variables. To proceed, we impose two restrictions: (1) $d(c)$ and $d_0(c)$ cannot rely on c ; and (2) $d = d(\text{no idea}) = d(c), \forall c > \bar{c}(h_{t+\epsilon-})$. In other words, the state variable remains constant until the next innovator who invests arrives, and conditional on such an arrival, the new state variable $d_0 + d_1$ does not depend on the new innovator's cost. The intuition for why the above restrictions are innocuous is as follows. Because the planner's value function V is concave in d , she would not let d change frequently. If no idea arrives, there is certainly no need to change d . If an idea arrives but is abandoned, it is as if no idea arrives and d does not respond either. When an innovator arrives and claims rights, the duration to prior innovators does need to change to make room for the new innovator. However, it is optimal for the planner to choose a constant d_0 , which would minimize the variation of the new state variable $d_0 + d_1$. We confirm this intuition by showing in Appendix B that the planner's objective would not improve even if we remove the two restrictions and allow for more general policies.

In the rest of this subsection, we shall simplify the PK constraint and the Bellman equation using the above restrictions. The next subsection solves for the optimal policies in the restricted problem.

The PK constraint and the Bellman equation in (8) and (9) rely on a small number $\epsilon > 0$, and they are essentially discrete-time approximations of their continuous-time counterparts. The standard approach to deriving the PK constraint and the Bellman equation in continuous time is to take limit $\epsilon \rightarrow 0$, but this requires a fair amount of algebra. We follow a more transparent approach below.

Recall that a new innovator invests if and only if his cost is below a cutoff, which equals d_1 . Such an innovator arrives with Poisson rate $\lambda F(d_1)$ and we denote his arrival time as τ . Under the restriction $d = d(\text{no idea}) = d(c)$, past innovators' duration promise remains constant from t to τ and jumps to d_0 at τ . Therefore, the PK constraint is

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

where $E \left[\int_t^\tau e^{-r(s-t)} ds \middle| h_t \right] y$ 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 $E \left[e^{-r(\tau-t)} \middle| h_t \right]$ brings d_0 , which is discounted to τ , back to time t . After some algebra, the PK constraint becomes⁷

$$d = \frac{1}{r + \lambda F(d_1)} y + \frac{\lambda F(d_1)}{r + \lambda F(d_1)} d_0. \quad (10)$$

Since $y \leq 1$, the PK constraint (10) is equivalent to an inequality⁸

$$d \leq \frac{1}{r + \lambda F(d_1)} + \frac{\lambda F(d_1)}{r + \lambda F(d_1)} d_0. \quad (11)$$

If (11) 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 innovator in order to satisfy the PK constraint (10). The Bellman equation for $V(d)$ is

$$V(d) = \max_{\substack{d_0 \geq 0, d_1 \geq 0, \\ d_0 + d_1 \leq r^{-1}}} E \left[e^{-r(\tau-t)} \middle| h_t \right] \left(\int_0^{d_1} (r^{-1} - c) \frac{f(c)}{F(d_1)} dc + V(d_1 + d_0) \right),$$

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

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

⁸Strictly speaking, the constraint $y \geq 0$ makes (10) more restrictive than (11). However, the constraint $y \geq 0$ never binds in our model and does not play any role.

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 . Simplifying the above equation yields

$$V(d) = \max_{\substack{d_0 \geq 0, d_1 \geq 0, \\ d_0 + d_1 \leq r^{-1}}} \frac{\lambda R(d_1)}{r + \lambda F(d_1)} + \frac{\lambda F(d_1)}{r + \lambda F(d_1)} V(d_1 + d_0), \quad (12)$$

subject to (11).

So far we have shown that the duration promises delivered by a mechanism $\{\bar{c}, I\}$ must satisfy the PK constraint (11). To fully justify the recursive approach, we also need to show that, if a sequence of duration promises satisfies the PK constraint, it can be delivered by some mechanism, i.e., there exists an indicator function I such that the duration promises are delivered by I . Lemma 9 in Appendix B shows this.

3.3 Optimal policies

We define sharing to be the allocation of rights to multiple innovators at the same history:

Definition 3 *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.

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

$$r\bar{d} + \lambda F(\bar{d})\bar{d} = 1. \quad (13)$$

The optimal policy rule is:

(i) if $d \in (\bar{d}, r^{-1}]$, then $d_0 + d_1 = d$ and the PK constraint (11) 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 $\lambda F(d_1)d_1$;

(ii) there is a number $d^* \in (0, \bar{d})$, such that if $d \in [d^*, \bar{d}]$, then $d_0 = 0$ and (11) binds,

$$\text{i.e., } d_1(d) = F^{-1}\left(\frac{1-rd}{\lambda d}\right) \equiv g(d);$$

(iii) if $d < d^*$, then $d_0 = 0$, $d_1(d) = g(d^*)$ and (11) is slack.

Variables d^* and \bar{d} are the cutoffs for the binding PK constraint and for the policy to exhibit sharing: the PK constraint binds if and only if $d > d^*$, and the policy function involves sharing if and only if $d > \bar{d}$. For small d , the PK constraint can be satisfied using only time before the next arrival, and therefore the PK constraint is slack; as a result the policy function involves no sharing since a slack PK constraint implies $d_0(d) = 0$.⁹ This immediately implies that $d^* \leq \bar{d}$.

In order to determine the values of d^* and \bar{d} we need to understand the optimal policy for $d > d^*$. The functions $g(d)$ and $h(1-rd)$ are the optimal $d_1(d)$ on $[d^*, \bar{d}]$ and $[\bar{d}, r^{-1}]$, respectively. In the following discussion, we will use \bar{d} to denote the value defined in (13), and then explain why the \bar{d} in (13) is indeed the cutoff above which sharing occurs. Both $g(d)$ and $h(1-rd)$ are decreasing in d . Monotonicity of $g(d)$ and $h(1-rd)$ reflects the fundamental scarcity in this model: a greater promise to incumbents reduces what can be offered to future innovators.

If $d = \bar{d}$, then $d_1 = g(\bar{d}) = \bar{d}$ and $d_0 = 0$, and so $d = \frac{1}{r+\lambda F(d_1)}$; the duration promise of $\bar{d} = d_1$ is exactly delivered by granting to the incumbents all instants before a new innovator whose cost is below \bar{d} arrives. This policy allocates every instant of time to some innovator, and perfectly smooths duration across innovators (including the incumbents) since the state variable remains \bar{d} and therefore the policy remains the same. \bar{d} is the highest promise for all innovators if the total duration of r^{-1} is to be 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} .

⁹By contradiction, suppose for some d the PK constraint is slack and $d_0(d) > 0$. The policy $(d_0(d), d_1(d))$ cannot be optimal, as it is dominated by $(\tilde{d}_0, \tilde{d}_1) = (d_0(d) - \epsilon, d_1(d) + \epsilon)$: the new policy $(\tilde{d}_0, \tilde{d}_1)$ still satisfies the PK constraint (because it is slack), increases the next innovator's investment cutoff (and welfare), and leaves the future state variable $d_0 + d_1$ unchanged.

This smoothing 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 allocate 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. Then $d_0 + d_1 = d$ ensures smoothing: the total duration promise stays at d forever and the cutoff \bar{c}_t equals $d_1 = h(1 - rd) < \bar{d}$ forever.¹⁰ Since both d_1 and d_0 are positive in this range, there is sharing.¹¹ Because there is no sharing at \bar{d} , but there is sharing for all $d > \bar{d}$, we conclude that the \bar{d} defined in (13) is indeed the cutoff \bar{d} where sharing begins.

When $d \in [d^*, \bar{d}]$, it follows from $d_0 = 0$ and the binding PK constraint that $d = \frac{1}{r + \lambda F(d_1)}$. Solving for d_1 yields $d_1 = F^{-1}\left(\frac{1-rd}{\lambda d}\right) \equiv g(d)$. Intuitively, $g(d)$ is the inverse function of the binding PK constraint. When $d \leq \bar{d}$, the PK constraint $d \leq \frac{1}{r + \lambda F(d_1)}$ can be rewritten as $d_1 \leq g(d)$. Appendix A defines $g(d^*)$ as the unconstrained optimal d_1 for the right-hand side of the Bellman equation under $d_0 = 0$. So the constraint $d_1 \leq g(d)$ binds only if $d > d^*$ (i.e., $g(d) < g(d^*)$); if $d \leq d^*$, d_1 equals the unconstrained maximizer $g(d^*)$.

Finally, we explain why $d^* < \bar{d}$, or equivalently, why the unconstrained maximizer of d_1 is above \bar{d} . The sequence of duration promises for future innovators is $(\bar{d}, \bar{d}, \bar{d}, \dots)$ if $d_1 = \bar{d}$, and is $(\bar{d} + \epsilon, h(1 - r(\bar{d} + \epsilon)), h(1 - r(\bar{d} + \epsilon)), \dots)$ if $d_1 = \bar{d} + \epsilon$ for a small $\epsilon > 0$. In both cases, all instants after the first innovator arrives are allocated (after the first innovator arrives, the PK constraint binds in both cases because the state variable stays forever at either \bar{d} or $\bar{d} + \epsilon$). Therefore, the total (discounted) duration promise for all innovators is equal to the (discounted) instants after the first innovator arrives, which is $\frac{\lambda F(d_1)}{r + \lambda F(d_1)}$. Raising d_1 from \bar{d} to $\bar{d} + \epsilon$ has a first-order benefit: if the first investing innovator arrives sooner, the instants after his arrival and the total duration promise will increase, which improves welfare. The downside is that it distorts future arrivals because

¹⁰Because $d_0 + d_1 = d$, 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 $d_1 = h(1 - rd)$. Because $d > \bar{d}$ and h is monotonic, $d_1 = h(1 - rd) < h(1 - r\bar{d}) = \bar{d}$.

¹¹That $d_0(d) > 0$ is because $\bar{d} - h(1 - r\bar{d}) = 0$ and $d - h(1 - rd)$ increases in d .

$d_1 > \bar{d}$ implies that the first innovator gets more duration than all subsequent ones, i.e., $\bar{d} + \epsilon > h(1 - r(\bar{d} + \epsilon))$. But for small ϵ this distortion cost is not first order, because the duration is nearly perfectly smoothed. Hence the planner will prefer $d_1 = \bar{d} + \epsilon$ 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 *From 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.*

Key to the sharing mechanism are two properties. First, the states with sharing are absorbing. If existing innovators will share duration with the next innovator, because the planner wants to perfectly smooth the duration promises to all future innovators, it is optimal for existing innovators to also share duration with every future innovator. Second, the planner prefers to enter the absorbing sharing states by offering the first innovator more duration than subsequent ones. As explained in the paragraph before Corollary 1, in the absence of old innovators (i.e., when the initial PK constraint is slack), doing so will increase the total duration promise for future innovators and hence will improve welfare.

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.¹² In Appendix B we discuss the

¹²Industries 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)).

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 is slack. The segment between d^* and \bar{d} is given 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 Implementing optimal duration with patents or licensing contracts

The previous section defines sharing to be the case where both the current and past innovators are promised some time selling the leading-edge product; the optimal contract employs such sharing, from (at the latest) the second innovation onward. In this section we describe how the optimal allocation of the previous section could be achieved through a system of patent rules, or alternatively through a licensing agreement entered into ex ante by the set of potential innovators.

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.

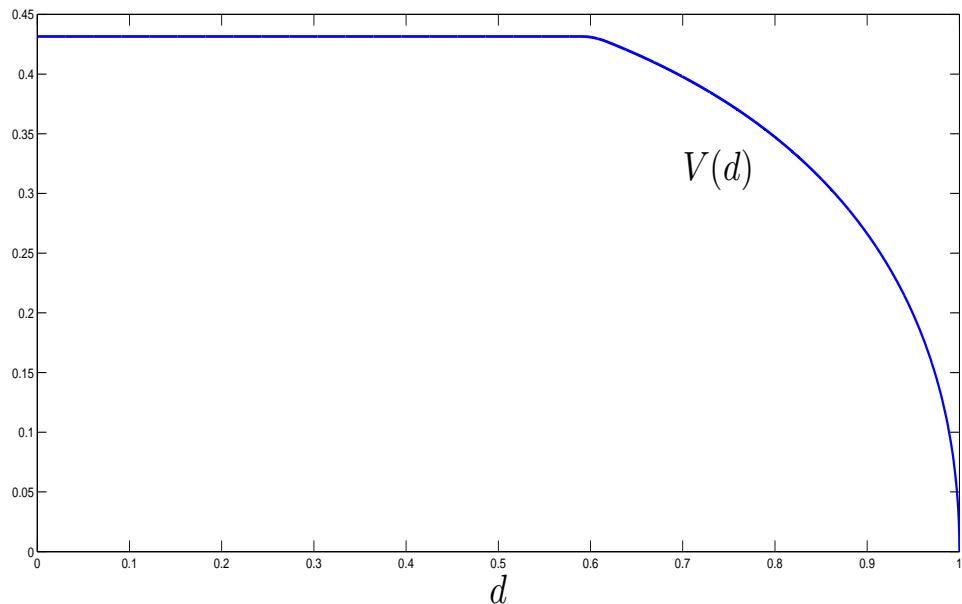


Figure 1: Value function with uniform density.

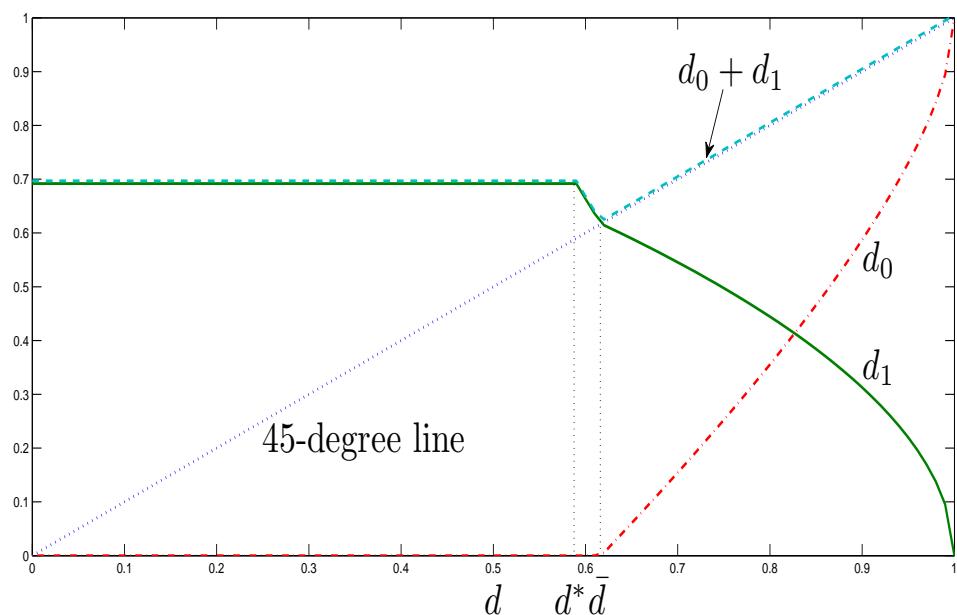


Figure 2: Policy functions with uniform density.

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 be achieved 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.

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 implement sharing through licensing contracts under the circumstance where later patents infringe on early ones. The licensing rules is 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 implementation, the innovators form an ever-growing 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 implementation 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.¹³

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 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.¹⁴ 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.¹⁵ The model proposed here

¹³Denote 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 $\sum_{n=1}^{\infty} \frac{d_1}{d_0+d_1}(1 - \frac{d_1}{d_0+d_1})^{n-1} \frac{1}{r+\lambda F(d_1)} \left(\frac{\lambda F(d_1)}{r+\lambda F(d_1)} \right)^{n-1} = \frac{\frac{d_1}{d_0+d_1} \frac{1}{r+\lambda F(d_1)}}{1 - \frac{\lambda F(d_1)}{r+\lambda F(d_1)} \frac{d_0}{d_0+d_1}} = \frac{\frac{1}{r+\lambda F(d_1)}}{d_0 + d_1 - \frac{\lambda F(d_1)}{r+\lambda F(d_1)} d_0} d_1 = d_1$, where the last equality follows from the binding PK constraint $d_0 + d_1 = \frac{1}{r+\lambda F(d_1)} + \frac{\lambda F(d_1)}{r+\lambda F(d_1)} d_0$.

¹⁴See, for example, Lerner and Tirole (2008).

¹⁵This idea is the basis of the model of patent pools in Lerner and Tirole (2004), where pools have

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. 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.

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. In practice, patent litigation is expensive, suggesting that shared rights may be costly. In this section we consider the extreme case where the costs of sharing are so high that sharing must be avoided completely.

The planner could avoid conflict resulting from shared rights 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 welfare consequences similar to the ones found in models of mergers, based on monopoly markup by the pool.

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 $d_1 \leq F^{-1}(\frac{1-rd}{\lambda d}) \equiv g(d)$ 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^{**} . Appendix A defines $g(d^{**})$ as the unconstrained optimal d_1 for the right-hand side of the Bellman equation under $d_0 = 0$.¹⁶ So the constraint $d_1 \leq g(d)$ binds only if $d > d^{**}$ (i.e., $g(d) < g(d^{**})$); if $d \leq d^{**}$, d_1 equals the unconstrained maximizer $g(d^{**})$.

Proposition 2 *When $d_0 = 0$ is imposed in (11) and (12), 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 $(0, \bar{d}]$ and \bar{d} is defined in (13) of Section 3. The rule can be summarized as $d_1 = \min(g(d), g(d^{**}))$. Furthermore, $d^{**} = \bar{d}$ if and only if $F(\bar{d}) \geq \bar{d}f(\bar{d})$.

Given $d^* < \bar{d}$, why is $d^{**} = \bar{d}$ possible? 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., $d^{**} = \bar{d}$).

¹⁶Although both $g(d^*)$ and $g(d^{**})$ are unconstrained optimal d_1 , $g(d^*) \neq g(d^{**})$ because on the right-hand side of the Bellman equation the value functions with and without sharing differ.

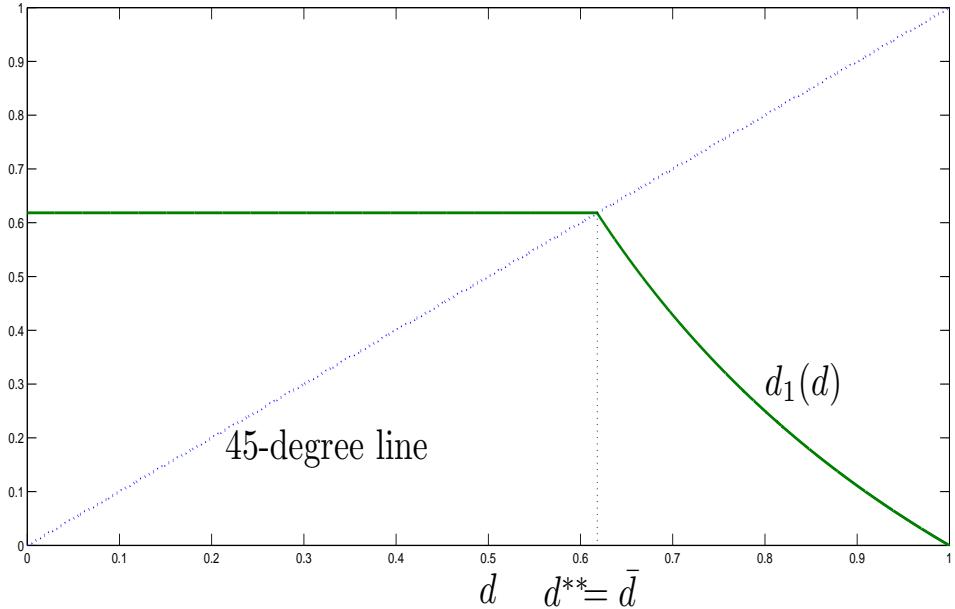


Figure 3: Conflict-free policy function with uniform density.

5.2 Dynamics without sharing

Denote the duration promise of the n th innovation as d_n (i.e., d_n equals $d_1(d_{n-1})$). The evolution of d_n critically depends on whether d^{**} equals \bar{d} . 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 for $d \leq \bar{d}$ is where the PK constraint is slack, while the segment for $d > \bar{d}$ is given 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.

Proposition 3 *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 $\lim_{n \rightarrow \infty} d_{2n} = g(d_\infty) \leq \bar{d}$.

When $d^{**} < \bar{d}$, the states fluctuate around \bar{d} because if $d_n > \bar{d}$, then $g(\bar{d}) = \bar{d}$ and the fact that $g(d)$ is decreasing imply $d_{n+1} = g(d_n) < g(\bar{d}) = \bar{d}$.¹⁷ This further implies that $d_{n+2} = \min(g(d_{n+1}), g(d^{**})) > g(\bar{d}) = \bar{d}$, as the monotonicity of $g(d)$ implies both $g(d_{n+1}) > g(\bar{d})$ and $g(d^{**}) > g(\bar{d})$. 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. In Appendix C we discuss the cycles in more detail.

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.

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} . (It can be verified that $h(1 - rd) > g(d)$ for $d > \bar{d}$.)

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

¹⁷That $g(d)$ decreases in d follows from the definition of $g(d) \equiv F^{-1}(\frac{1-rd}{\lambda d})$. That $g(\bar{d}) = F^{-1}(\frac{1-r\bar{d}}{\lambda \bar{d}}) = \bar{d}$ follows from the definition of \bar{d} , $r\bar{d} + \lambda F(\bar{d})\bar{d} = 1$.

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 no-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, the planner implements 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 $d^* < d^{**}$.¹⁸

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, suppose the contract enters a perpetual cycle for all initial promises $d > \bar{d}$ and the perpetual cycle is independent of d . (Appendix C contains such an example.) 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

¹⁸Although we could not prove $d^* < d^{**}$ analytically, it holds in all of our numerical examples.

$\lambda F(\bar{d})$. If d 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 property in Corollary 4 that no-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. One interpretation of this pioneer is as an “standard-essential patent,” which is commonly associated with key early innovations in a patent pool.

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 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. This high promise matches the privileged position of a standard-essential patent in practice. 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 satisfies $g(d_p) = \frac{1}{r+\lambda F(r^{-1})}$, i.e., $g(g(d_p)) = r^{-1}$. 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 $g(d_p) = \frac{1}{r+\lambda F(r^{-1})}$ 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, $g(d_p)$, satisfies $g(g(d_p)) = r^{-1}$, the 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 (and costly) path of progress without sharing.

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 implemented with 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

Proof of Lemma 1: Inequalities (5) and (7) imply $U(c, 0) = 0$. ■

Proof of Lemma 2: If $C(h_t) = 1$, then (6) implies that $U(c, c, 1) \geq 0$. If an innovator has cost $c' < c$, reports c , and invests, then his payoff is positive because

$$U(c', c, 1) = U(c, c, 1) + c - c' \geq c - c' > 0.$$

By contradiction, suppose $C(h'_t) = 0$. Then inequality (7) implies

$$0 = U(c', 0) \geq U(c', c, 1),$$

which contradicts $U(c', c, 1) > 0$. This contradiction implies $C(h'_t) = 1$. ■

Proof of Lemma 3: Because $U(c, 0) = 0$ in σ , the modified mechanism $\tilde{\sigma}$ delivers the same payoff to innovator i . To see that $\tilde{\sigma}$ is still IC, note that if another innovator has cost c' but reports c , then his payoffs when he does and does not invest under $\tilde{\sigma}$ are $U(c', c, 1) = -c'$ and $U(c, 0) = 0$, respectively. IC constraints (6) and (7) are still satisfied. To show $V(\tilde{\sigma}) = V(\sigma)$, note that the cutoff \bar{c} in $\tilde{\sigma}$ is identical to that in σ , and the cutoff is what matters for welfare. ■

Proof of Lemma 4: The definition of $d_1(h_t)$ and (5) imply

$$U(c, c, 1) = U(c, 0) + d_1(h_t) - c \leq d_1(h_t) - c.$$

If $c = \bar{c}$, then $0 \leq U(c, c, 1)$ implies $d_1(h_t) \geq c = \bar{c}$. If $c < \bar{c}$, pick a small $\epsilon > 0$ such that $c < \bar{c} - \epsilon$. We have

$$\begin{aligned} U(c, c, 1) &\geq U(c, \bar{c} - \epsilon, 1) \\ &= U(\bar{c} - \epsilon, \bar{c} - \epsilon, 1) + \bar{c} - \epsilon - c \\ &\geq \bar{c} - \epsilon - c. \end{aligned}$$

Therefore, $d_1(h_t) \geq \bar{c} - \epsilon$ for any small $\epsilon > 0$. This shows that $d_1(h_t) \geq \bar{c}$. ■

Proof of Lemma 5: If $c \leq \bar{c}$, define $\{\tilde{I}(h_s, i); s \geq t, h_s \text{ follows } h_t\}$ such that

$$E \left[\int_t^\infty e^{-r(s-t)} \tilde{I}(h_s, i) ds \middle| h_t \right] = \bar{c}.$$

Because $E \left[\int_t^\infty e^{-r(s-t)} I(h_s, i) ds \middle| h_t \right] \geq \bar{c}$, \tilde{I} can be obtained from I by changing some ones in $\{I(h_s, i); s \geq t, h_s \text{ follows } h_t\}$ to zeros.

Define the fees $\{\tilde{\phi}(h_s) : s \geq t, h_s \text{ follows } h_t, \tilde{I}(h_s, i) = 1\}$ such that

$$U(c, 0) \equiv E \left[\int_t^\infty e^{-r(s-t)} \tilde{I}(h_s, i) (n_s - 1 - \tilde{\phi}(h_s)) ds \middle| h_t \right] = 0. \quad (14)$$

It is easy to verify that $\tilde{\sigma}$ is IC. Under $\tilde{\sigma}$, (5) holds because $U(\tilde{c}, 0) = 0, \forall \tilde{c}$. Inequality (6) holds because if $C(h_t) = 1$ and $\tilde{c} \leq \bar{c}$, then $U(c, c, 1) = \bar{c} - c = U(c, \tilde{c}, 1)$. Inequality (7) holds because if $C(h_t) = 0$ and $\tilde{c} \leq \bar{c}$, then $U(c, 0) = 0 > \bar{c} - c = U(c, \tilde{c}, 1)$.

To show $V(\tilde{\sigma}) = V(\sigma)$, note that the cutoff \bar{c} in $\tilde{\sigma}$ is identical to that in σ , and the cutoff is what matters for welfare. ■

Proof of Lemma 6: Suppose by contradiction $V(\tilde{h}_s) < V(h_t)$. Let $\{I(h_x, i) : t < t_i \leq x\}$ be the optimal indicator functions for future innovators who arrive after history h_t . Define the set of unassigned instants as

$$S \equiv \{h_x : t \leq x, I(h_x, j) = 0 \text{ for all } j \text{ such that } t < t_j \leq x\}.$$

The discounted expected duration of all the instants in S is equal to $d(h_t) = d(\tilde{h}_s)$.

Starting from history \tilde{h}_s , we design an indicator function \tilde{I} that achieves the same value as $V(h_t)$. First, \tilde{I} treats future innovators after history \tilde{h}_s in the same way as I treats future innovators after history h_t . Second, the planner uses the instants from S to fulfill the promises made to prior innovators (those who arrive before history \tilde{h}_s). If there is only one prior innovator at \tilde{h}_s , that innovator is entitled to the entire $d(\tilde{h}_s)$ and hence will receive all the instants in S . If there are two prior innovators who are entitled to durations D_1 and D_2 ($D_1 + D_2 = d(\tilde{h}_s)$), respectively, then the planner can split S into two disjoint subsets, S_1 and S_2 ($S_1 \cup S_2 = S$), and assign the instants in S_1 to the first innovator and the instants in S_2 to the second. Note that S_1 and S_2 must be adjusted carefully to ensure that the durations (D_1, D_2) are exactly delivered. Because S contains a continuum of instants and each instant has a measure of zero, this split is always feasible.¹⁹ Similar arguments hold when more than two prior innovators are entitled to a share of d . To summarize, we can design an indicator function \tilde{I} that achieves the same value as $V(h_t)$, contradicting that $V(\tilde{h}_s)$ is the optimal value starting from history \tilde{h}_s . ■

For convenience, we transform the Bellman equation (12) into the following (15). 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

¹⁹In fact, there is a continuum of different ways to choose the right subsets S_1 and S_2 . This means that the optimal indicator function in the planner's problem is not unique.

can rewrite (12) as

$$V(d) = \max_{\substack{\tilde{d} \in [(r+\lambda)^{-1}, r^{-1}] \\ d_0 \in [0, r^{-1} - g(\tilde{d})]}} \tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V\left(g(\tilde{d}) + d_0\right) \quad (15)$$

subject to $d \leq \tilde{d} + (1 - r\tilde{d})d_0$.

Proposition 1 *The solution to (15) is*

$$V(d) = \begin{cases} d^*\lambda R(g(d^*)) + (1 - rd^*)V_{\mathbb{R}}(g(d^*)), & d \leq d^*; \\ d\lambda R(g(d)) + (1 - rd)V_{\mathbb{R}}(g(d)), & d \in [d^*, \bar{d}]; \\ V_{\mathbb{R}}(d) \equiv r^{-1}\lambda R(h(1 - rd)), & d \geq \bar{d}, \end{cases} \quad (16)$$

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

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

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

$$\left. \left(\tilde{d}\lambda R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d})) \right)' \right|_{\tilde{d}=\bar{d}} = V'_{\mathbb{R}}(\bar{d}) < 0. \quad (17)$$

Third, $V'(\cdot)$ is continuous at \bar{d} , which implies that $V(\cdot)$ is globally concave. It follows from (17) that

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

Fourth, we verify the Bellman equation (15). Because V coincides with $V_{\mathbb{R}}$ when $d \geq \bar{d}$, we only verify (15) on $[0, \bar{d}]$. Pick a feasible (d_0, \tilde{d}) such that $d_0 > 0$. We will show that $V(d) \geq \tilde{\lambda}R(g(\tilde{d})) + (1 - r\tilde{d})V_{\mathbb{R}}(g(\tilde{d}) + d_0)$ as follows. If $\tilde{d} \geq \bar{d}$, then

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

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

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

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 $V(\cdot)$. ■

When $d_0 = 0$ is imposed in (15), the Bellman equation becomes

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

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

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

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 $\lambda\tilde{d}R(g(\tilde{d})) + (1 - r\tilde{d})w(g(\tilde{d}))$ 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 $\tilde{\lambda}R(g(\tilde{d})) + (1 - r\tilde{d})W(g(\tilde{d}))$. 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.

Fourth, we show that $d^{**} = \bar{d}$ if and only if $F(\bar{d}) \geq \bar{d}f(\bar{d})$.

Necessity: If $d^{**} = \bar{d}$, then

$$W(d) = \begin{cases} r^{-1}\lambda R(\bar{d}), & d \leq \bar{d}; \\ d\lambda R(g(d)) + (1 - rd)W(\bar{d}), & d > \bar{d}. \end{cases} \quad (18)$$

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

$$\begin{aligned} \max_{d \in [(r+\lambda)^{-1}, \bar{d}]} d\lambda R(g(d)) + (1 - rd)W(g(d)) &= \max_{d \in [(r+\lambda)^{-1}, \bar{d}]} d\lambda R(g(d)) + (1 - rd)\Gamma(g(d)) \\ &= \max_{d \in [(r+\lambda)^{-1}, \bar{d}]} \Omega(d), \end{aligned}$$

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 (18) satisfies the Bellman equation $W = TW$.

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

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

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 3: 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 $d_1 \geq d_3$. \blacksquare

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 $d \in [0, r^{-1}]$.

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

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

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 $dR(g(d))$. \blacksquare

Lemma A.2 If $v(\cdot)$ is decreasing and concave, then $(1 - rd)v(g(d))$ is concave in $d \in [0, r^{-1}]$ 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. \blacksquare

Lemma A.3 (Relaxed problem) When $d_0 \geq 0$ is not imposed, the solution to (12) is

$$V_{\mathbb{R}}(d) = r^{-1}\lambda R(h(1 - rd)), \quad (19)$$

which is concave and strictly decreasing in $d \in [0, r^{-1}]$. 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. Because d_1 decreases in d , $V'_{\mathbb{R}}(d)$ is decreasing in d .

Next we verify the Bellman equation (12). 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 $r\tilde{d}_0 = 1 - r\tilde{d}_1 - \lambda\tilde{d}_2 F(\tilde{d}_2)$ into the PK constraint yields

$$\frac{r}{r + \lambda F(\tilde{d}_1)}\lambda F(\tilde{d}_1)\tilde{d}_1 + \frac{\lambda F(\tilde{d}_1)}{r + \lambda F(\tilde{d}_1)}\lambda F(\tilde{d}_2)\tilde{d}_2 \leq 1 - rd. \quad (20)$$

The objective on the right side of (12) 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 (20). This verifies the Bellman equation. ■

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

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

■

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

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

■

Appendix B: Extensions

Allowing value to vary with cost

Rather than taking the size of quality improvement 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 an IC mechanism 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 Lemma 1 shows that there is no reward or punishment.

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 IC 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 IC 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 IC 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

$$V(d) = \max_{d_0, d_1, e_0, e_1} \quad \frac{(1+r)e_0}{r} - \frac{e_0^2}{2\theta_0} + \frac{1}{1+r} \left(\frac{(1+r)e_1}{r} - \frac{e_1^2}{2\theta_1} \right)$$

subject to $e_0 = \theta_0 \left(1 + \frac{d_0}{1+r} \right), \quad e_1 = \theta_1 d_1, \quad d_0 + d_1 = 1 + \frac{\bar{d}_0}{1+r}.$

In this problem, the optimal sharing rule depends only on $\frac{\theta_1}{\theta_0}$ (see Figure 4):

- (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.

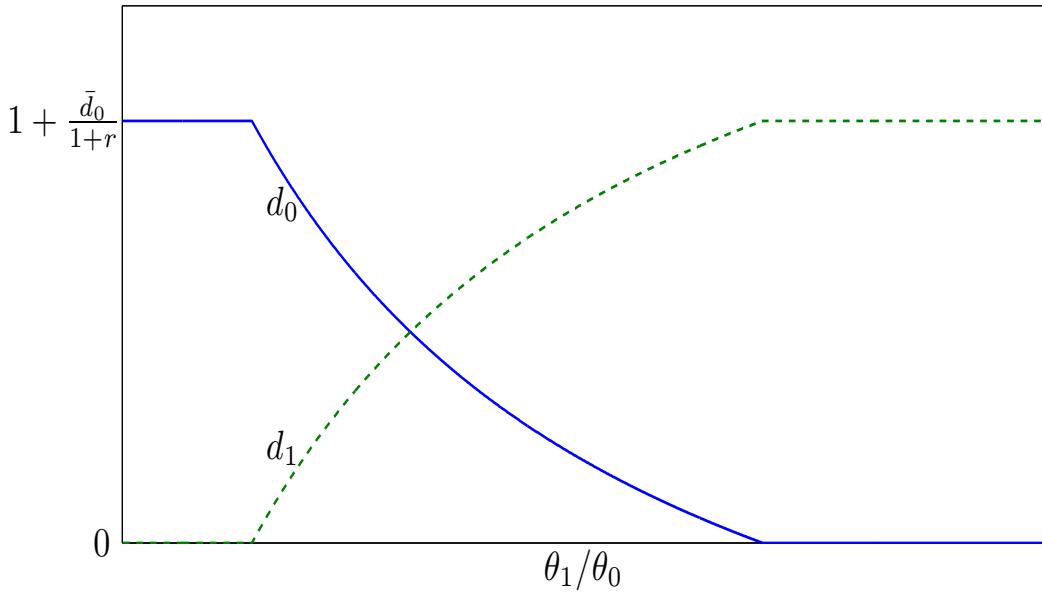


Figure 4: Policy functions with endogenous improvement size.

Optimality of type-independent policies

Recall from subsection 3.1 that $d_0(c)$ and $d(c)$ are the duration promises for the incumbents when a new type- c innovator receives a recommendation that he should and

should not invest, respectively. We show below that the planner cannot improve welfare by allowing $d_0(c)$ or $d(c)$ to depend on c . In particular, consider a type-independent policy in which the incumbents receive $d_0 \equiv \frac{\int_0^{\bar{c}} d_0(c)f(c)dc}{F(\bar{c})}$ for all $c \leq \bar{c}$ and $d(\text{not invest}) \equiv \frac{\int_{\bar{c}}^{\infty} d(c)f(c)dc}{1-F(\bar{c})}$ for all $c > \bar{c}$.

Lemma 7 *The type-independent policy achieves a weakly higher social welfare.*

Proof: The planner's expected continuation value after the innovation is $\int_0^{\bar{c}} V(d_0(c) + d_1)f(c)dc$, which is lower than the value under policy (d_0, d_1) because

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

where the inequality follows from the concavity of $V(\cdot)$. Using a similar proof, we can show that

$$\int_{\bar{c}}^{\infty} V(d(c))f(c)dc \leq (1 - F(\bar{c}))V(d(\text{not invest})).$$

■

Optimality of $d = d(\text{no idea}) = d(c)$

Recall from subsection 3.1 that $d(\text{no idea})$ and $d(c)$ are the duration promises for the incumbents when no idea arrives at time $t + \epsilon$ and when a new innovator arrives at time $t + \epsilon$ with a recommendation that he should not invest, respectively. Since we have shown that $d(c)$ does not depend on c , we denote $d(c)$ as $d(\text{not invest})$.

The promise-keeping constraint in continuous time is

$$rd = y + \lambda F(d_1)(d_0 - d) + \lambda(1 - F(d_1))(d(\text{not invest}) - d) + \dot{d}, \quad (21)$$

where $\dot{d} \equiv \lim_{\epsilon \downarrow 0} \frac{d(\text{no idea}) - d}{\epsilon}$ is the time derivative of the state variable before the next idea arrives. To understand (21), it would be illuminating to first consider the case of $\lambda = 0$, that is, the case with no arrival of new innovators. Rewrite (21) 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 $F(d_1)(d_0 - d) + (1 - F(d_1))(d(\text{not invest}) - d)$ 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

$$\dot{d} = rd - y - (\lambda F(d_1)(d_0 - d) + \lambda(1 - F(d_1))(d(\text{not invest}) - d)).$$

The dynamic programming problem is

$$\begin{aligned} rV(d) = \max_{d_0, d_1, d(\text{not invest}), y, \dot{d}} \quad & \lambda R(d_1) + \lambda F(d_1)(V(d_1 + d_0) - V(d)) \\ & + \lambda(1 - F(d_1))(V(d(\text{not invest})) - V(d)) + V'(d)\dot{d}, \\ \text{subject to} \quad & (21). \end{aligned}$$

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 8 *If $V(\cdot)$ is concave, then $d(\text{not invest}) = d$ and $\dot{d} = 0$. That $\dot{d} = 0$ implies $d = d(\text{no idea})$.*

Proof: Let $\mu(d)$ be the Lagrange multiplier on the PK constraint (21). The first-order conditions for $d(\text{not invest})$ and \dot{d} are, respectively,

$$\lambda(1 - F(d_1))V'(d(\text{not invest})) + \lambda(1 - F(d_1))\mu(d) = 0, \quad V'(d) + \mu(d) = 0,$$

which imply that $V'(d(\text{not invest})) = V'(d)$, and hence $d(\text{not invest}) = d$. The envelope condition is

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

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

Equivalence between the sequence problem and the recursive problem

Lemma 9 *Suppose the planner faces state variable d at time t_0 , i.e., d is the remaining promise for an old innovator with index 0. Let t_i be the arrival time of the i th innovator ($i \geq 1$), and $d_1(t_i)$ and $d_0(t_i)$ be the duration promises to the i th and earlier innovators at t_i , respectively. If $(d_0(t_i), d_1(t_i))_{i=1}^{\infty}$ satisfies the PK constraint in the recursive problem, i.e.,*

$$d_0(t_i) + d_1(t_i) \leq \frac{1}{r + \lambda F(d_1(t_{i+1}))} + \frac{\lambda F(d_1(t_{i+1}))}{r + \lambda F(d_1(t_{i+1}))} d_0(t_{i+1}), \quad (22)$$

then there exists an indicator function I that delivers promise $d_1(t_j)$ to any innovator j :

$$d_1(t_j) = E \left[\int_{t_j}^{\infty} e^{-r(t-t_j)} I(h_t, j) dt | h_{t_j} \right], \quad \forall j \geq 0. \quad (23)$$

(When $j = 0$, we interpret $(d_0(t_j), d_1(t_j))$ as $(0, d)$.)

Proof: At t_i , define for any innovator $j \in \{0, 1, 2, \dots, i\}$

$$D_j(t_i) \equiv \begin{cases} d_1(t_i), & j = i; \\ \frac{d_1(t_j)}{d_0(t_j) + d_1(t_j)} \left(\prod_{k=j+1}^{i-1} \frac{d_0(t_k)}{d_0(t_k) + d_1(t_k)} \right) d_0(t_i), & j < i. \end{cases}$$

(Intuitively, $D_j(t_i)$ is innovator j 's remaining duration at time t_i delivered by the indicator function I to be constructed below.) In the time interval $[t_i, t_{i+1})$, construct $I(h_t, j)$ for each $j \in \{0, 1, 2, \dots, i\}$ such that

$$E \left[\int_{t_i}^{t_{i+1}} e^{-r(t-t_i)} I(h_t, j) dt | h_{t_i} \right] = D_j(t_i) - E [e^{-r(t_{i+1}-t_i)} | h_{t_i}] D_j(t_{i+1}). \quad (24)$$

This construction is feasible, because (1) the total duration in $[t_i, t_{i+1})$ is more than the total duration assigned to all innovators;²⁰ and (2) the interval $[t_i, t_{i+1})$ contains a continuum of instants and each instant has a measure of zero.

We verify (23) in the following. Equation (24) states

$$D_j(t_j) = E \left[\int_{t_j}^{t_{j+1}} e^{-r(t-t_j)} I(h_t, j) dt | h_{t_j} \right] + E \left[e^{-r(t_{j+1}-t_j)} | h_{t_j} \right] D_j(t_{j+1}).$$

Applying (24) to $D_j(t_{j+1})$ yields

$$\begin{aligned} D_j(t_j) &= E \left[\int_{t_j}^{t_{j+1}} e^{-r(t-t_j)} I(h_t, j) dt | h_{t_j} \right] + E \left[e^{-r(t_{j+1}-t_j)} | h_{t_j} \right]. \\ &\quad \left(E \left[\int_{t_{j+1}}^{t_{j+2}} e^{-r(t-t_{j+1})} I(h_t, j) dt | h_{t_{j+1}} \right] + E \left[e^{-r(t_{j+2}-t_{j+1})} | h_{t_{j+1}} \right] D_j(t_{j+2}) \right) \\ &= E \left[\int_{t_j}^{t_{j+2}} e^{-r(t-t_j)} I(h_t, j) dt | h_{t_j} \right] + E \left[e^{-r(t_{j+2}-t_j)} | h_{t_j} \right] D_j(t_{j+2}). \end{aligned}$$

Repeating the above procedure, we have

$$D_j(t_j) = E \left[\int_{t_j}^{t_{j+k}} e^{-r(t-t_j)} I(h_t, j) dt | h_{t_j} \right] + E \left[e^{-r(t_{j+k}-t_j)} | h_{t_j} \right] D_j(t_{j+k}), \quad \forall k \geq 1,$$

which yields (23) after we take limit $k \rightarrow \infty$. ■

Appendix C: Further details for Section 5

The two-period cycle in Proposition 3 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 5 and 6 plot two dynamics of duration promises starting with $d_1 = 0.76$. We provide sufficient conditions for the two cases.

Proposition 4 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 $d_\infty = g(d^{**})$.

²⁰

$$\begin{aligned} \sum_{j=0}^i \left(D_j(t_i) - E \left[e^{-r(t_{i+1}-t_i)} | h_{t_i} \right] D_j(t_{i+1}) \right) &= d_0(t_i) + d_1(t_i) - \frac{\lambda F(d_1(t_{i+1}))}{r + \lambda F(d_1(t_{i+1}))} d_0(t_{i+1}) \\ &\leq \frac{1}{r + \lambda F(d_1(t_{i+1}))} = E \left[\int_{t_i}^{t_{i+1}} e^{-r(t-t_i)} dt | h_{t_i} \right], \end{aligned}$$

where the first equality follows from the definition of D_j and the inequality follows from (22).

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

Proof:

- (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 $d > \bar{d}$.

■

Proposition 5 Let $b \in (\bar{d}, r^{-1})$ be the unique solution of

$$\lambda \int_0^b (b - c)f(c)dc + rb - 1 - \lambda R(h(1 - rb)) = 0. \quad (25)$$

- (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 $c \in (\bar{d}, b]$.

Proof: 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 (25). 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 4, 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

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

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

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

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

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

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 $\alpha^2(r\check{d})^\alpha(1-r\check{d})$ remains bounded, $1 - \frac{x}{\alpha} < r\check{d}$ 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 $c \in (\bar{d}, b]$.

■

We end 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 7 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 4 is satisfied and cycles last forever. However, when α is sufficiently large, the sufficient condition in Proposition 5 is satisfied and the states converge to \bar{d} . Figure 8 plots the policy function when $\alpha = 6$: in contrast to that in Figure 7, $d_1(d_1(\cdot))$ is flatter than the 45-degree line and cycles disappear eventually.

References

- Akcigit, U., Kerr, W. R., 2010. Growth through heterogeneous innovations. Harvard Business School Entrepreneurial Management Working Paper No. 11-044.
- Chari, V., Golosov, M., Tsyvinski, A., 2012. Prizes and patents: Using market signals to provide incentives for innovations. Journal of Economic Theory 147 (2), 781–801.
- 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 (1), 20–33.
- Henry, E., 2010. Promising the right prize. Working paper, London School of Business.
- 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.

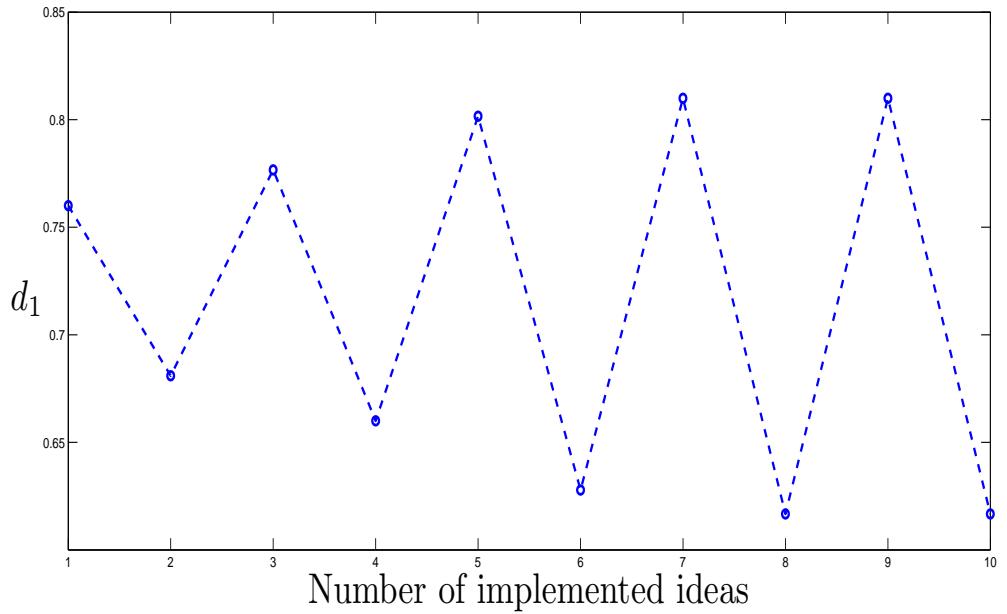


Figure 5: Perpetual fluctuation with power density $f(c) = 3c^2$.

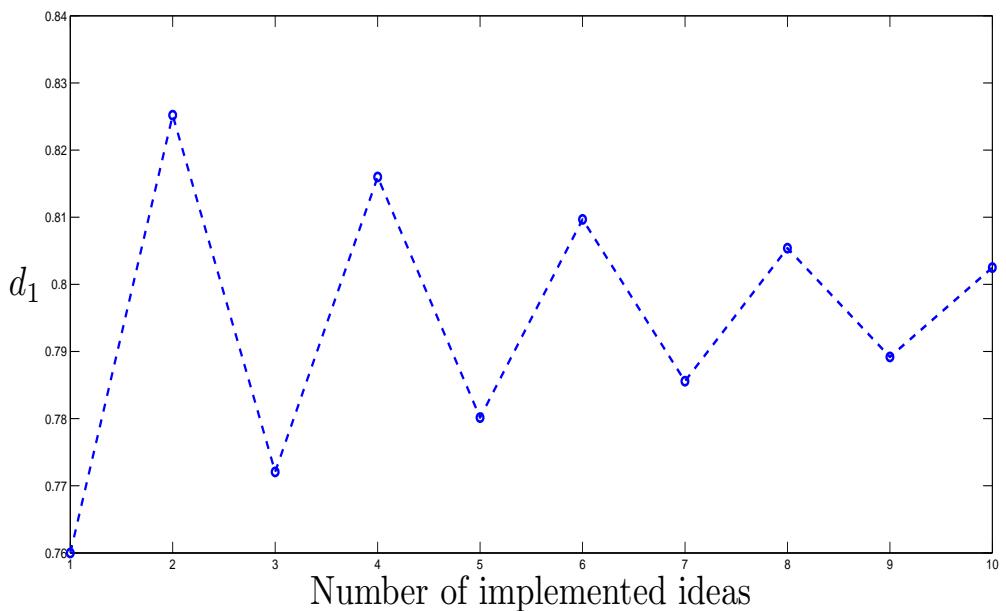


Figure 6: Degeneration with power density $f(c) = 6c^5$.

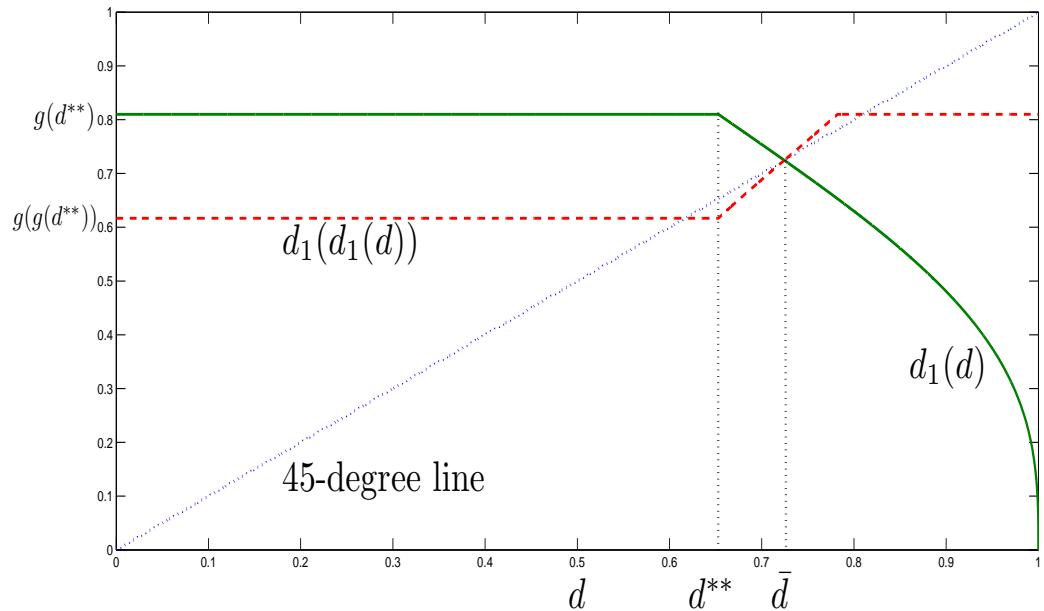


Figure 7: Policy function with power density $f(c) = 3c^2$.

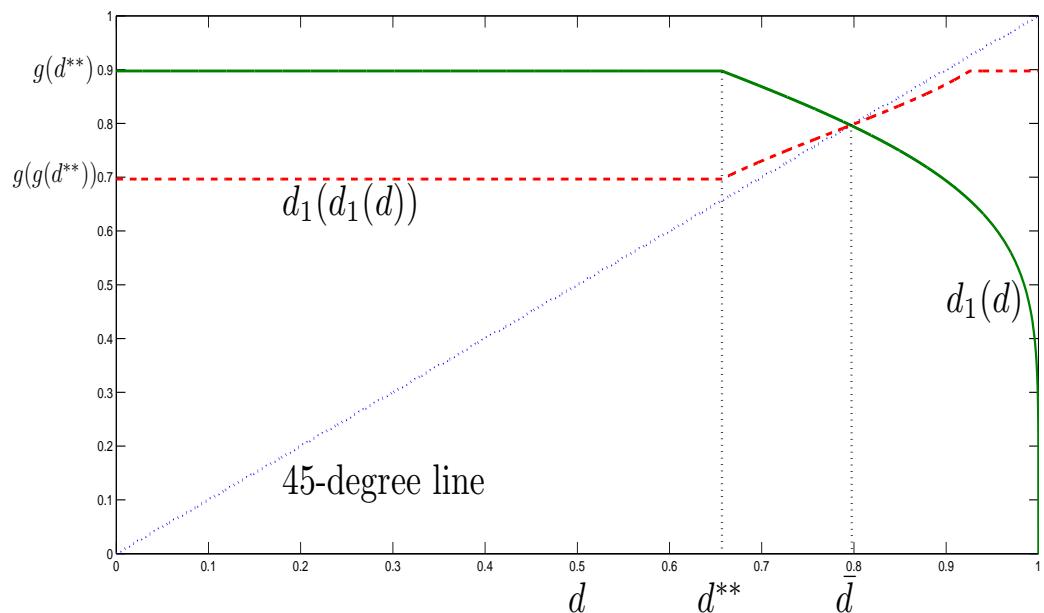


Figure 8: Policy function with power density $f(c) = 6c^5$.

- 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., 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., Bessen, J. E., 2005. Lessons for patent policy from empirical research on patent litigation. *Lewis & Clark Law Review* 9 (1), 1–27.
- Mill, J. S., 1883. *Principles of political economy: with some of their applications to social philosophy*. D. Appleton and Co.
- 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.
- 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., 2012. Market power screens willingness-to-pay. *The Quarterly Journal of Economics* 127 (4), 1971–2003.