Patentability, RD direction, and cumulative innovation

Chen, Yongmin and Pan, Shiyuan and Zhang, Tianle

University of Colorado Boulder, Zhejiang University, Lingnan University

August 2016

Patentability, R&D Direction, and Cumulative Innovation

Yongmin Chen* Shiyuan Pan† Tianle Zhang‡

Revised June 2017

Abstract: We present a model of cumulative innovation where firms can conduct R&D in both a safe and a risky direction to produce quality improvements over the current technology. As patentability standards rise, an innovation in the risky direction is less likely to receive a patent, which decreases the static incentive for new entrants to conduct risky R&D, but increases their dynamic incentive because of the longer duration—and hence higher reward—for incumbency. These, together with a strategic substitution and a market structure effect, result in an inverted-U shape in the risky direction but a U shape in the safe direction for the relationship between R&D intensity and patentability standards. There exists a patentability standard that induces the efficient innovation direction, whereas R&D is biased towards (against) the risky direction under lower (higher) standards. The optimal patentability standard generally distorts the R&D direction in order to increase the industry innovation rate.

Keywords: cumulative innovation, patentability standards, R&D intensity, R&D direction, rate of innovation, innovation direction

*University of Colorado Boulder, USA; yongmin.chen@colorado.edu

†Zhejiang University, China; shiyuanpan@zju.edu.cn

‡Lingnan University, Hong Kong; tianlezhang@ln.edu.hk

**Corresponding to: Shiyuan Pan, School of Economics, Zhejiang University, Hangzhou, China; shiyuanpan@zju.edu.cn. We thank two referees, an editor, and participants of several conferences for helpful comments. Shiyuan Pan gratefully acknowledges financial support from the Key Project of the National Social Science Foundation of China (No. 15AJY003) and the Project of the Scientific Research Foundation for the Returned Overseas Chinese Scholars of the Ministry of Education of China. Tianle Zhang gratefully acknowledges the Hong Kong Research Grant Council for research support under Grant# 23501215.

1 Introduction

A central issue in the economics of innovation is how patent policy may affect innovative activities. The recent literature has examined this issue in the context of *cumulative* innovation, where discoveries build on each other, under a standard assumption that firms pursue innovations along a single research direction. In many industries, however, firms can conduct R&D in multiple directions to achieve a specific goal, as, for example, the development of a next generation color copier in the early 1990s by Fuji Xerox, of a new mobile system by Ericsson in the mid-1990s, and of an X Terminal workstation by the Hewlett Packard in the late 1980s (Birkinshaw and Lingblad, 2001).[1] The purpose of this paper is to inquire how patent policy, specifically patentability standards, may affect the rate and direction of cumulative innovation in an industry where firms can conduct R&D in multiple directions.

We consider a situation where there are two research directions, A and B, for a sequence of innovations (or new products) that deliver higher product qualities over time. The quality improvement of an innovation in direction B may range stochastically from low to high while that in direction A is at some intermediate level. Hence, if an innovation is patentable only when its quality improvement (or innovation size) is sufficiently large, as for instance implied by the requirement of a minimum inventive step, there will be a range of quality thresholds, or patentability standards (S), under which innovations in direction A are always patentable but an innovation in direction B may not. We will focus on patentability standards in the interior of such a range, and call A the safe direction while B the risky direction.[2]

If innovation is a one-time activity that ends with the successful introduction of a new product, a (marginally) higher patentability standard would discourage R&D in the risky direction by making it harder to obtain a patent and the rents associated with it through

[1] In particular, facing the possible emergence of a third-generation mobile system with high bandwidth, Ericsson in the mid-1990s funded research teams to separately develop two different standards: a more radical new technology called WCDMA and a new standard called EDGE upgrading the existing technologies.

[2] That is, we consider patentability standards that are high enough to exceed the left tail of innovations in the risky direction but not so high to make a safe innovation not patentable. We rule out by assumption situations where patentability standards are so high that a safe innovation is not patentable but a risky one can (in which case the problem would be reduced to analyzing innovations only in the risky direction).

this direction, which we shall call the *threshold* effect, whereas it would have no impact on R&D in the safe direction, provided that there are no (dis)economies of scope in R&D and that the return to a successful patentable discovery in one direction is not diminished by that in the other. In this static setting, a higher S reduces industry R&D through the threshold effect, and it also allocates relatively more resources in the safe direction than in the risky direction, which can reduce the expected size of innovation if a successful innovation through the risky direction has a higher expected quality improvement than that through the safe direction.

The issue is more complex if innovations are cumulative, as we assume in this paper. Specifically, we consider the following model that builds on and extends Hunt (2004) by having two research directions: Suppose that $n + 1$ firms have entered an industry. At any time, one of them is the leader and the other n firms are challengers. The challengers are in a patent race to develop a new product that improves upon the current leader's. When a challenger succeeds in a patentable innovation, it becomes the new leader to replace the current one, who then joins the rank of challengers; and this process repeats itself indefinitely. In this dynamic setting, a marginal increase in the patentability standard will increase the value of being a leader because it will take longer before the leader is replaced by a successful challenger. This *incumbency-prolonging* effect can potentially increase the incentive for R&D in both innovation directions, even though the threshold effect from a higher S will still have a negative impact on the incentive for R&D in the risky direction.[3]

Moreover, the changes in the R&D incentives in the two different directions will interact with each other, giving rise to a dynamic *strategic substitution effect* between the two directions: When the R&D intensity in one direction becomes higher (or lower), it exerts an opposite force on the R&D intensity in the other direction. In particular, an increase of

[3] It has been found in the literature that innovation and competition may have a non-monotonic relationship because, while more intense competition may lower rents from a single innovation, it could increase innovation incentives under cumulative innovation due to the dynamic effect and the desire to "escape competition" (Aghion et al., 2005) The incumbency-prolonging effect is also due to the dynamic effect, but it works through the channel of patentability standards.

R&D in one direction induces the next innovation discovery to come sooner, which lowers the profit from incumbency and thus reduces the incentive for R&D in the other direction. This turns out to be the crucial force that leads to new effects of patentability standards under multiple research directions.

Finally, as we shall assume, a firm needs to incur a fixed cost to enter the market in order to conduct R&D and innovate. Therefore, patentability standards, by impacting the expected return to R&D in each direction, also affects the number of entrants in the free entry equilibrium. Our analysis will examine how this *market-structure* effect interacts with the other forces in the model.

We find that as patentability standards rise, R&D intensity in the risky direction first rises and then falls, exhibiting an inverted-U shape, whereas R&D intensity in the safe direction is U-shaped, initially decreasing and then increasing. Thus, the incumbency-prolonging effect is the dominating force in the risky direction when S is low, but it is dominated by the negative threshold effect when S is high. More surprising is that despite the positive impact from the incumbency-prolonging effect, increases in S initially lower R&D in the safe direction, due to the strategic substitution effect.

We also find that as S increases, the industry rate of innovation initially goes up and eventually falls down, reaching its maximum at some intermediate level. The market-structure effect plays a balancing role: there will be more firms when the expected return from R&D investment is higher, which moderates the effects of patentability standards on R&D intensities both for each firm and for the industry.

We further compare the market equilibrium with the solutions that maximize social welfare. First, in relation to the first-best innovation rate, we show that R&D intensities and the number of entrants in the free entry equilibrium are deficient. This is due to the familiar intuition that a firm's private innovation incentive does not internalize the positive externalities to consumers.[4] Second, compared to the first-best innovation direction, we find

[4] In our model, consumers benefit from a non-patentable innovation immediately. For a patentable innovation, the increase in equillibirum price offsets the consumer gain from the quality improvement during the

that there exists a critical value of patentability standard, $\hat{S}$, such that the equilibrium R&D direction coincides with the first best when $S = \hat{S}$, and it is biased towards (against) the risky direction when S is below (above) $\hat{S}$.[5] For the second-best social welfare maximization problem, in which a hypothetical social planner can only set the patentability standard but not the R&D and entry activities of firms, the optimal S balances the three goals of moving towards the socially optimal innovation rate, towards the socially optimal innovation direction, and towards the socially optimal market structure. Thus, in general, the second best patentability standard will be different from $\hat{S}$, from the S that maximizes the number of innovating firms, and from the S that maximizes the rate of innovation either for an individual firm or for the industry.

Our paper is related to the existing theoretical literature on patents and cumulative innovation, which has studied models with R&D along a single direction and offered mixed findings on the effects of patent protection. For example, O'Donoghue (1998) and O'Donoghue et al. (1998) suggest that stronger patent protection has positive effects on the rate of innovation, provided that ex-ante agreement or contracting between innovators is efficient, whereas Bessen and Maskin (2009) and Segal and Whinston (2007) find cases where the effects are negative. Horowitz and Lai (1996) consider a model in which longer patents increase the size but decrease the frequency of the innovation. They show that the patent length that maximizes the rate of innovation is finite (or intermediate).[6] As we mentioned earlier, our model is most closely related to Hunt (2004), who studies patentability and

period, but the consumer gain is realized once the next innovation replaces the current one so that the new price only reflects the new quality improvement. Thus consumers benefit from a patentable innovoation not immediately but dynamically. Notice that, under competition, there is also a business-stealing effect that potentially results in excessive R&D and entry. In our model, as in Hunt (2004), the positive externality dominates.

[5] Intuitively, when S is low, innovations in the risky direction are patentable even when the quality improvement is small, which motives socially excessive R&D in the risky direction, relative to the safe direction. And the opposite is true when S is high. Since we measure innovation or R&D direction by the ratio of R&D intensities in the two directions, R&D can be efficient in both directions and yet biased towards one direction.

[6] Chen et al. (2014) find that stronger patent protection can affect cumulative innovation either positively or negatively, and the effect is generally non-monotonic. Empirically, some recent studies on cumulative innovation (Murray et al., 2007; Furman and Stern, 2011; Galasso and Schankerman, 2013; Williams, 2013; Sampat and Williams, 2014) find no evidence of a relationship.

cumulative innovation in a model with R&D only in one direction that corresponds to the risky direction in our paper. By allowing multiple R&D directions, we introduce the important strategic substitution effect and offer several new insights. In particular, in contrast to the result in Hunt that the patentability standard affects innovation only through a market structure effect, with no impact on each innovating firm's R&D intensity, we show that it also affects innovation through its impact on R&D intensities, in ways that are non-monotonic and somewhat unexpected. Thus, in our model, patentability standards affect industry innovation through both the extensive margin (number of entrants) and the intensive margin (R&D intensities). Moreover, our results on innovation (or R&D) direction are novel in this literature.

Our paper is also related to a large literature, broadly defined as on R&D portfolio and the direction of innovation. Earlier studies have focused on the issue of how competition may affect the choice between safe and risky research projects for a stand-alone innovation.[7] Some authors have found, under the assumption of winner-take-all, that competition leads to over-investment in risky R&D projects because it magnifies the negative externality of investment by one firm on other firms' probability to win the patent.[8] Others, however, have argued that investment in risky R&D project decreases with the strength of competition, because the negative externality of the risky R&D becomes small when competition strengthens, if each firm pursues multiple patents (Cabral, 1994; Kwon, 2010).[9] Recent studies have examined sequential innovation. Acemoglu (2011) considers a model with sequential innovation and multiple research paths but only one research path is commercially

[7] Acemoglu (2002) argues that profit incentives may shape the direction of technical change and therefore determine the equilibrium bias of technology.

[8] In a classic paper, Dasgupta and Maskin (1987) show that in R&D races firms select a too high expected rate of technological changes which, in most cases, induces excessively risky research projects (see also Bhattacharya and Mookherjee, 1986; Klette and de Meza, 1986).

[9] Relatedly, Anderson and Cabral (2007) study a game where firms choose the variance of a stochastic innovation outcome. They find that the level of equilibrium variance may be greater, smaller, or equal to the social optimum. Aghion, Dewatripont and Stein (2008) provide a framework for evaluating the advantages and disadvantages of academic research as opposed to private-sector research and show that it is possible for ideas to be privatized sooner than is socially optimal. Choi and Gerlach (2014) study the R&D choice between easy and difficult projects that are complementary for the production of a final product. They find that firms tend to invest excessively on the easy innovation due to hold-up problems.

active at any point in time. He shows that the possibility of changing preferences can induce inefficiency because the returns from innovation are only realized for those generations where the research line is commercially active. More recently, Hopenhayn and Squintani (2016) investigate the incentive to innovate among multiple directions in a growth model, finding that the equilibrium allocation of researchers across R&D lines is suboptimal, with too many pursuing "hot" R&D lines. Bryan and Lemus (2016) consider a directional model where firms both race toward easy projects and do not fully appropriate the value of their inventions.[10]

In the rest of the paper, we describe our model and its equilibrium in Section 2. In Section 3, we establish our results on how the patentability standard affects the rates of innovation, as measured by the R&D intensities of each firm in the two directions and by the overall R&D intensity of the industry, and how it affects the direction of innovation, as measured by the ratio of the innovation rates in the two directions. We also discuss to what extent each of the main features of our model—particularly multi-periods, uncertainty, multiple research direction, and reward sizes—is responsible for these effects. Section 4 contains our welfare results, comparing the equilibrium rate and direction of innovation with the social optimum, and discussing optimal patentability policy as the second best. In Section 5, we discuss how our findings might be affected if we relax the assumption that incumbents do not engage in R&D or if patentability standards are state-dependent. Section 6 concludes. Our main results are illustrated through a numerical example in Appendix A, and proofs that are more technical in nature are relegated to Appendix B. Technical details analyzing the extension with incumbent innovation are contained in Appendix C.

[10] They also show that patens can distort effort towards either incremental or radical innovation and, similarly as we do, find that the second-best patent policy necessarily induces inefficiency. Notice that while our paper shares the common interest of the aforementioned two papers in considering innovation directions, we analyze a rather different model and focus on the role of patentability standard, with insights on how it impacts R&D directions and intensities, as well as on its optimal design.

2 The Model

Time is continuous and is divided into periods, $t = 0, 1, 2, ...,$ between stochastic discoveries by innovating firms. There are $n+1$ firms in the industry, one of whom is the incumbent and the others are challengers in each period. At period t, the incumbent, through a patented innovation at an earlier period, can produce a product that has quality q_t. Each of the challengers conducts R&D to further improve the product quality.

There are two possible research directions for the challengers, direction A—the safe direction, and direction B—the risky direction. A successful innovation through direction A will result in a certain quality improvement, Δ_A.[11] A successful innovation through direction B will yield an uncertain quality improvement, Δ_B, which is a random variable with cumulative distribution function $G(\cdot)$ and continuous density $g(\cdot)$ on support $\left[0, \overline{\Delta}_B\right]$. As we pointed out before, this formulation closely follows Hunt (2004), with the main difference being that he considers R&D only along a single uncertain direction corresponding to B here.

A challenger decides on a R&D portfolio by choosing the R&D intensity in each of the two directions. We assume that each innovation occurs according to a Poisson process. The cost for a challenger to maintain an arrival rate λ_z in research direction $z \in \{A, B\}$ is $C(\lambda_z)$,[12] which is strictly increasing and twice continuously differentiable, with $C(0) = C'(0) = 0$, $C''(\cdot) > 0$, and $\lim_{\lambda_z \to \infty} C'(\lambda_z) = \infty$.[13] We shall also refer to λ_z as the R&D intensity in direction z.

[11] We can allow Δ_A to be stochastic, provided that its variance is sufficiently small.

[12] Our formulation implicitly assumes that a firm's total R&D costs are $C(\lambda_A) + C(\lambda_B)$, separable across directions. We thus consider situations where R&D inputs are not substitutable between the two innovation directions, possibly because—for instance—they require researchers who specialize in different technologies. This is a restrictive assumption. It would be more realistic to allow the substition of R&D inputs between alternative research directions, but the analysis could be much more complicated. We leave this for possible future research.

[13] We follow Lee and Wilde (1980) in assuming that innovation is produced through flow costs, which, as they point out, may generate additional innovation as firms enter, relative to the case where fixed costs are required for innovation (Loury, 1979). Notice that we allow the "corner" case where each firm chooses to conduct R&D only in one direction. Under our assumptions on the cost function, however, the equilibrium will be interior.

The statutory life of a patent is assumed to be infinite, even though the patent life effectively ends when the next patentable invention occurs. To be awarded a patent, the quality improvement from an invention needs to meet a minimum improvement size, or the patentability standard, S. In practice, the patentability standard (or requirement) can correspond to the requirement of non-obviousness in the American patent code, or of the inventive step in Europe. For the purpose of this paper, we assume that $S \in \left[0, \bar{\Delta}_B\right]$ and $S < \Delta_A$. Thus, an innovation achieved through the safe direction is always patentable, whereas $\theta(S) \equiv 1 - G(S)$ is the probability that an innovation in the risky direction is granted a patent. When an innovation is not protected by a patent, it becomes freely available to the public, in which case we assume that competition drives the profit from marketing the product to zero. Notice that the more stringent the patentability requirement, other things equal, the smaller the probability that the challenger can profitably market her innovation achieved through the risky direction.

We assume that at the beginning of period $t = 0$, there is a large number of firms, each deciding whether to pay a one-time fixed investment cost k to enter the market. Thus, the number of challengers, n, is endogenously determined by the free-entry condition. If a challenger wins the race for a patentable innovation, it becomes the incumbent in the next period, and the previous incumbent becomes a challenger. If a challenger succeeds in an innovation that does not meet the patentability standard, then the incumbent maintains its leader position, and all $n + 1$ firms enter into a new period of patent race. The innovation arrival rates and the costs to achieve them remain the same after any discovery, whether patentable or not. Therefore, in either case, the relative positions of the $n + 1$ firms in the market are the same, and hence the choice problem for any firm in the market is stationary. We denote the discount rate, common for all firms, by r.

The market contains a representative consumer, who demands one unit of the product per period. The consumer's valuation for a product is equal to its quality. The marginal cost of production for any firm is normalized to zero. The incumbent and the challengers engage in price competition. Thus, when the incumbent's product quality exceeds the next closest quality by Δ, its flow profit is exactly Δ until the arrival of a new patentable innovation. The challengers earn no flow profit.

As in Hunt (2004) and other studies in this literature, we shall focus on an equilibrium where only challengers, but not the incumbent, will invest in R&D. Incumbents tend to have lower incentive to invest in R&D than entrants due to their existing profit. The assumption that they make no investment is more extreme, and it is made mainly for analytical tractability. [We discuss in Section 5 how our analysis could be extended to a setting where incumbents also engage in R&D.] Notice that in our model, players rotate their roles as the incumbent and the challengers over time, so a firm may only temporarily stop investing. In our analysis that follows, by construction, the strategies by the challengers and the incumbent will constitute a stationary Markov Perfect Equilibrium (MPE).[14]

We shall maintain the following assumption throughout the paper:

$$0 < rk < \Delta_A \leq E\left[\Delta_B\right] \equiv \int_0^{\overline{\Delta}_B} \Delta_B dG\left(\Delta_B\right), \tag{A1}$$

because of two considerations: First, we wish to ensure that a positive number of firms will be willing to enter the market to pursue innovation in each direction, which will require $0 < rk < \Delta_A$. Second, we are interested especially in situations where a successful innovation in the risky direction yields a higher expected quality improvement, which is captured by $\Delta_A < E\left[\Delta_B\right]$, but we also allow $\Delta_A = E\left[\Delta_B\right]$ in order to isolate the effect of uncertainty.

If a challenger innovates through the safe direction, she becomes an incumbent and receives a profit flow of

$$\pi_A = \Delta_A \tag{1}$$

until she is replaced by a future challenger. If the challenger succeeds in the risky direction,

[14] There could potentially be another equilibrium where an incumbent from direction B may conduct R&D if the realized $\Delta_B > S$ is relatively small, but its analysis appears to be untractable.

the expected profit flow (conditional on the innovation being patentable) is

$$\pi_B = \frac{1}{\theta(S)} \int_S^{\overline{\Delta}_B} \Delta_B dG(\Delta_B). \tag{2}$$

Notice that

$$\frac{\partial \pi_B}{\partial S} = \frac{-Sg(S)\theta(S) + g(S)\int_S^{\overline{\Delta}_B} \Delta_B dG(\Delta_B)}{[\theta(S)]^2} = \frac{g(S)}{\theta(S)}(\pi_B - S) \geq 0,$$

where $\pi_B \geq S$ because

$$\int_S^{\overline{\Delta}_B} \Delta_B dG(\Delta_B) \geq \int_S^{\overline{\Delta}_B} S dG(\Delta_B) = \theta(S) S.$$

It follows that

$$\pi_B \equiv \pi_B(S) \geq \pi_B(0) = E[\Delta_B].$$

This, together with (1) and assumption (A1), implies that $\pi_B \geq \pi_A > rk > 0$. Thus, entry to pursue innovation in each direction can be profitable. The equilibrium number of entrants in the market will be determined simultaneously as the arrival rate of innovation in each direction, as we show next.

Because all challengers are symmetric, we focus on stationary equilibria where they choose identical R&D strategies. Specifically, at such a stationary MPE, which is assumed to exist uniquely, let V_z^I be the value of being an incumbent through type-z innovation and V^E the value of being a challenger, all of which are evaluated at the beginning of a period. Then V_A^I, V_B^I and V^E satisfy the following Hamilton-Jacobi-Bellman equations[15]:

$$rV_A^I = \pi_A + n(\lambda_A + \theta\lambda_B)\left(V^E - V_A^I\right), \tag{3}$$

[15] Notice that the probability that any two innovations succeed simultaneously is zero. Because we are constructing an equilibrium in which the incumbent does not invest, no matter what its realized quality improvement is, the value function V_B^I below is not contingent on the realization of Δ_B.

$$rV_B^I = \pi_B + n\left(\lambda_A + \theta\lambda_B\right)\left(V^E - V_B^I\right), \tag{4}$$

and

$$rV^E = \lambda_A\left(V_A^I - V^E\right) - C_A\left(\lambda_A\right) + \theta\lambda_B\left(V_B^I - V^E\right) - C_B\left(\lambda_B\right). \tag{5}$$

Equations (3), (4) and (5) suggest that the value of being an incumbent depends on the type of innovation that has led to the incumbency.[16]

From (5), the challenger chooses optimal λ_A and λ_B, which respectively satisfy the first-order conditions:[17]

$$C_A'\left(\lambda_A\right) = V_A^I - V^E, \tag{6}$$

and

$$C_B'\left(\lambda_B\right) = \theta\left(V_B^I - V^E\right). \tag{7}$$

The free entry condition implies

$$V^E = k. \tag{8}$$

From (3), (6) and (8), we find

$$V_A^I - V^E = \frac{\pi_A - rk}{r + n(\lambda_A + \theta\lambda_B)} = C_A'\left(\lambda_A\right). \tag{9}$$

Similarly, from (4), (7) and (8), we have

$$V_B^I - V^E = \frac{\pi_B - rk}{r + n(\lambda_A + \theta\lambda_B)} = \frac{C_B'\left(\lambda_B\right)}{\theta}. \tag{10}$$

Substituting (9) and (10) into (5) yields

$$\lambda_A C_A'\left(\lambda_A\right) + \lambda_B C_B'\left(\lambda_B\right) - C_A\left(\lambda_A\right) - C_B\left(\lambda_B\right) - rk = 0. \tag{11}$$

[16] Note that the size of quality improvement appears with the corresponding probability, even though the incumbent knows its exact value after innovation is successful. Hence π_B is shown in the right hand of (4). We note again that in general there may be other equilibria, possibly with asymmetric R&D by entrants. We focus on the specific equillibrium by assumption.

[17] The properties of the cost functions ensure that the second-order conditions are satisfied.

The system of equations, (9), (10) and (11), determine the three equilibrium values λ_A^*, λ_B^* and n^*. In particular, from (9) and (10), the equilibrium number of challengers can be expressed as

$$n^* = \left[\frac{\pi_A - rk}{C_A'(\lambda_A^*)} - r\right] \cdot \frac{1}{\lambda_A^* + \theta\lambda_B^*} = \left[\frac{\theta(\pi_B - rk)}{C_B'(\lambda_B^*)} - r\right] \cdot \frac{1}{\lambda_A^* + \theta\lambda_B^*}. \tag{12}$$

We illustrate the equilibrium of the model with an example in Appendix A.

3 The Rates and Direction of Innovation

We are now in a position to examine how the patentability standard, S, may affect the rates and direction of innovation. We first consider the effects of S on the equilibrium R&D intensities, λ_A^* and λ_B^*, which can be viewed as each entrant's innovation rates in the safe and risky directions, respectively. Recall that λ_A^*, λ_B^* and n^* are determined by (9), (10) and (11). In the appendix, we show the following by using the Cramer's rule:

$$\frac{\partial\lambda_A^*}{\partial S} = \frac{g(S)\lambda_B^*(\lambda_A^* + \theta\lambda_B^*)\left[C_A'(\lambda_A^*)\right]^2 C_B''(\lambda_B^*)(rk - S)}{|M|} \tag{13}$$

and

$$\frac{\partial\lambda_B^*}{\partial S} = \frac{-g(S)\lambda_A^*(\lambda_A^* + \theta\lambda_B^*)\left[C_A'(\lambda_A^*)\right]^2 C_A''(\lambda_A^*)(rk - S)}{|M|}, \tag{14}$$

where

$$|M| = -(\lambda_A^* + \theta\lambda_B^*)C'(\lambda_B^*)\, C''(\lambda_A^*)\, C''(\lambda_B^*)\left[\lambda_A^*(\pi_A - rk) + \theta\lambda_B^*(\pi_B - rk)\right] < 0$$

since $\pi_B \geq \pi_A > rk$. Thus, if $S < rk$, then $\frac{\partial\lambda_A^*}{\partial S} < 0$ and $\frac{\partial\lambda_B^*}{\partial S} > 0$; while if $S > rk$, then $\frac{\partial\lambda_A^*}{\partial S} > 0$ and $\frac{\partial\lambda_B^*}{\partial S} < 0$. This leads to the following result, where we define

$$d(S) \equiv \frac{\lambda_B^*}{\lambda_A^*} \tag{15}$$

as the innovation direction.

Proposition 1 *As S increases, λ_B^* first increases and then decreases, whereas λ_A^* first decreases and then increases, reaching the maximum and the minimum, respectively, at $S = rk$. Moreover, innovation direction $d(S)$ has an inverted-U shape, maximized at $S = rk$.*

Proof. See Appendix B. ■

Interestingly, R&D intensities in both directions vary non-monotonically with S, in contrast to the result in Hunt (2004) that R&D intensity is invariant with the patentability standard. As we discussed in the introduction, a marginal increase in S has both a threshold and an incumbency-prolonging effect: it reduces the probability of obtaining a patent in the risky direction but increases the value of being an incumbent; the former can be attributed to the uncertain size of risky innovations, while the latter is due to the multi-periods feature of our model. The initial increase of λ_B^* in S is driven by the incumbency-prolonging and strategic substitution effects, which outweigh the threshold effect, whereas the latter effect dominates so that λ_B^* decrease in S when $S > rk$. Another crucial force is the strategic substitution effect due to the presence of two research directions: When the R&D intensity in one direction becomes lower (higher), it positively (negatively) impacts the R&D intensity in the other direction due to (the reverse of) the incumbency-prolonging effect. The interactions of these three effects are subtle, and together they jointly determine how λ_B^* and λ_A^* vary with S.[18]

Thus, in the free entry equilibrium of our model, patentability standards impact industry R&D not only through the number of firms (the extensive margin), but also through changes in the R&D intensities in different directions (the intensive margin). Notice that in (14), if $\lambda_A^* = 0$, then λ_B^* would be independent of S, and our results would coincide with Hunt's.[19]

[18] Proposition 1, as well as our other results to follow, holds for $E[\Delta_B] \geq \Delta_A$. Thus, our main results do not depend on whether the expected innovation size is the same in the two diretions or is higher in direction B. Notice also that our assumption on the strict convexity of $C(\cdot)$, which implies a motive for diversification in R&D directions, is important for the unique existence of interior λ_A^* and λ_B^*, and hence it is also important for the strategic substitution effect.

[19] See the equation after (A.7) on pp. 421 in Hunt (2004), except that, in Hunt, (i) there is an industry-specific productivity parameter, which we assume to be 1; and (ii) the reservation value of the product is

In our model, innovation direction, $d(S)$, is measured by each challenger's R&D intensity in the risky direction relative to that in the safe direction, which determines the relative rates of innovation achieved through the two directions. The inverted-U shaped $d(S)$, with its maximum attained at $S = rk$, follows directly from the shapes of $\lambda_B^*(S)$ and $\lambda_A^*(S)$. Since the expected size of each innovation is weakly higher in the risky direction than in safe direction, one might think that it would be desirable to choose $S = rk$. However, the overall expected innovation rate of each challenger,

$$\rho \equiv \lambda_A^* \Delta_A + \lambda_B^* E[\Delta_B] = \lambda_A^* \Delta_A \left(1 + d(S) \frac{E[\Delta_B]}{\Delta_A}\right),$$

depends also on how different $\lambda_B^*(S)$ and $\lambda_A^*(S)$ are. Hence, ρ may not be maximized at rk.

For the industry innovation rate, we need to further consider the number of entrants in equilibrium (n^*), which is also a function of S. The equilibrium overall innovation rate of the industry can be defined as:

$$R \equiv n^* \rho. \tag{16}$$

The result below indicates that the shape of $R \equiv R(S)$ is consistent with that of $d(S)$.

Proposition 2 *As S rises, R initially increases and eventually decreases, reaching its maximum when S is at some intermediate level.*

Proof. See Appendix B. ■

As in Hunt (2004), the industry rate of innovation (R) is maximized when the patentability standard is neither too high nor too low. However, the channels through which S affects R differ in the two models. In Hunt, as S increases, the equilibrium number of firms to conduct R&D in the market first increases and then decreases, whereas the equilibrium R&D intensity remains unchanged. Our model entails a second channel: the changes in the R&D

the level of its quality multiplied by p, and we assume $p = 1$.

intensities, since

$$\frac{\partial R}{\partial S}=n^{*}\frac{\partial \rho}{\partial S}+\rho\frac{\partial n^{*}}{\partial S},$$

and in our model, λ_A^* and λ_B^*—and hence also ρ—in general vary with S.

Define S_R as the patentability requirement that maximizes the innovation rate of the industry $R=R(S)$:

$$S_R=\arg\max\{R(S)\}.$$

In Appendix B, we show that

$$\frac{\partial R}{\partial S}|_{S=rk}>0, \tag{17}$$

which immediately leads to:

Remark 1 *If R is a single-peaked function of S, then $S_R>rk$.*

Therefore, the patentability standard that maximizes the overall rate of innovation in the industry is higher than $S=rk$, which maximizes λ_B, provided that $R(S)$ is single-peaked.[20]

Results in this section and the next section will be illustrated in Example 1 of Appendix A.

4 Efficient Innovation Incentive and Optimal S

In this section, we compare the equilibrium and the efficient incentives for cumulative innovation, where "efficient" means welfare-maximizing or the first-best; and we study how to choose S optimally at the market equilibrium. Specifically, we seek to answer two questions. First, if one could directly choose the number of entrants and the R&D intensities to maximize social welfare (i.e., the first best), what would be these choices and how would they differ from those in the market equilibrium? Second, if policy can choose patentability

[20] The single-peak condition on R is needed in Remark 1, because Proposition 2 does not rule out the possibility that R has multiple peaks as S increases on $[0,\Delta_A)$. Notice also that we assume $S<\Delta_A$. If S were to be higher than Δ_A, then no innovation would happen in direction A and the result in Hunt (2004) would directly imply $S_R>rk$.

standards, but not firms' innovative activities, what should be the optimal S (i.e., the second best)? Subsection 4.1 addresses the two questions in terms of the values for λ_A, λ_B, and n, while subsection 4.2 considers the questions from the perspective of innovation directions.

4.1 Comparing R&D Intensities and the Number of Entrants

When there are n challengers, each choosing R&D intensities λ_A and λ_B in safe and risky directions, respectively, total welfare is

$$W = \frac{n}{r}\left[\lambda_A \frac{\Delta_A}{r} - C_A(\lambda_A) + \lambda_B \frac{E[\Delta_B]}{r} - C_B(\lambda_B) - rk\right], \tag{18}$$

where $\frac{\Delta_A}{r}$ and $\frac{E[\Delta_B]}{r}$ are the expected social values of innovations generated by one innovating firm through the safe and risky directions, respectively.[21] The expression inside the square brackets in (18) is thus the instantaneous social benefit from one innovating firm, and there are n independent innovating firms for the industry, multiplied by $\frac{1}{r}$ to account for the discounted sum of the instantaneous benefits.

At the first best where a hypothetical social planner directly chooses λ_A, λ_B and n to maximize W, the welfare-maximizing λ_A^o and λ_B^o satisfy the following first-order conditions:

$$C'(\lambda_A^o) = \frac{\Delta_A}{r} \qquad \text{and} \qquad C'(\lambda_B^o) = \frac{E[\Delta_B]}{r}. \tag{19}$$

Notice that the efficient R&D intensities equate their marginal social benefits and costs. Comparing (19) to (9) and (10) and noticing that

$$\theta(\pi_B - rk) = \int_S^{\overline{\Delta}_B} \Delta_B dG(\Delta_B) - \theta rk < E[\Delta_B], \tag{20}$$

we find that the efficient R&D intensities are higher than those in the free entry equilibrium:

[21] Our model is one of quality ladders, where the benefit of a quality improvement to the society lasts forever. Thus, the expected social value of an innovation through the safe direction (discounted to the moment of its discovery) is $\int_0^\infty e^{-rt}\Delta_A dt = \frac{\Delta_A}{r}$. Similarly, the expected social value of an innovation through the risky direction is $\int_0^\infty e^{-rt} E(\Delta_B)\, dt = \frac{E(\Delta_B)}{r}$.

$\lambda_z^o > \lambda_z^*$, for $z = A, B$. Intuitively, the quality improvement from an innovation benefits the society permanently, but the innovating firm can capture the rents only before it is replaced by the next innovation. Moreover, some quality improvements along direction B are not patentable, which further lowers a firm's innovation incentive below the efficient level.

Moreover, since $\lambda_z C_z'(\lambda_z) - C(\lambda_z)$ increases in λ_z and $\lambda_z^o > \lambda_z^*$ for $z = A, B$, utilizing (19) and (11), we have

$$\begin{aligned} & \lambda_A^o \frac{\Delta_A}{r} - C_A(\lambda_A^o) + \lambda_B^o \frac{E[\Delta_B]}{r} - C_B(\lambda_B^o) - rk \\ = \; & \lambda_A^o C_A'(\lambda_A^o) - C_A(\lambda_A^o) + \lambda_B^o C_B'(\lambda_B^o) - C_B(\lambda_B^o) - rk \\ > \; & \lambda_A^* C_A'(\lambda_A^*) - C_A(\lambda_A^*) + \lambda_B^* C_B'(\lambda_B^*) - C_B(\lambda_B^*) - rk = 0. \end{aligned} \tag{21}$$

Hence, as in Hunt (2004), the efficient number of firms is $n^o = \infty > n^*$.

Summarizing the discussions above, we have:

Proposition 3 *Compared to the first-best, R&D intensities and the number of entrants are deficient under the free entry equilibrium.*

We note two related points. First, the result that aggregate R&D intensity is higher under the social optimum than under the market equilibrium does not rely on the number of firms at the first-best being infinite. Second, the key to the result of deficient entry is that the expected social benefit of adding one more firm exceeds the sum of its entry and R&D costs. Our analysis has not considered other mechanisms, such as increasing—rather than constant—marginal cost of production, that can also lead to deficient entry.

When policy can choose the patentability standard whereas firms choose R&D intensities under free entry to maximize their private benefits, the optimal choice of S is also called the second-best problem. Let $W(S)$ be the welfare in equilibrium at the second best. Then, from (18),

$$\frac{\partial W(S)}{\partial S} = \frac{1}{r}\frac{\partial n^*}{\partial S}\left[\lambda_A^*\frac{\Delta_A}{r} - C_A(\lambda_A^*) + \lambda_B^*\frac{E[\Delta_B]}{r} - C_B(\lambda_B^*) - rk\right] + \frac{n^*}{r}\left\{\left[\frac{\Delta_A}{r} - C_A'(\lambda_A^*)\right]\frac{\partial \lambda_A^*}{\partial S} + \left[\frac{E[\Delta_B]}{r} - C_B'(\lambda_B^*)\right]\frac{\partial \lambda_B^*}{\partial S}\right\}. \quad (22)$$

The optimal patentability standard, denoted by S^*, coincides with the one that maximizes the number of entrants in Hunt (2004) if $\lambda_A^* \equiv 0$. To see this, notice that if $\lambda_A^* \equiv 0$, our model reduces to that in Hunt (2004), implying $\frac{\partial \lambda_A^*}{\partial S} = 0$, and by (14) $\frac{\partial \lambda_B^*}{\partial S} = 0$. Hence, $\frac{\partial W(S)}{\partial S} = 0$ implies $\frac{\partial n^*}{\partial S} = 0$. However, in our model

$$\left\{\left[\frac{\Delta_A}{r} - C_A'(\lambda_A^*)\right]\frac{\partial \lambda_A^*}{\partial S} + \left[\frac{E[\Delta_B]}{r} - C_B'(\lambda_B^*)\right]\frac{\partial \lambda_B^*}{\partial S}\right\}$$

is generally not zero when $\frac{\partial n^*}{\partial S} = 0$, and thus the optimal patentability standard generally differs from the one that maximizes n^*.

From Remark 1, the S that maximizes the industry innovation rate (R) exceeds rk, provided that R is a single-peaked function of S. If W is a single-peaked function of S, then S^* also exceeds rk. To see this, note that, from (9), (10) and (20),

$$C_A'(\lambda_A^*) < \frac{\Delta_A}{r} \quad \text{and} \quad C_B'(\lambda_B^*) < \frac{\theta(\pi_B - rk)}{r} < \frac{E[\Delta_B]}{r}.$$

Thus, noticing $\frac{\partial \lambda_A^*}{\partial S}|_{S=rk} = \frac{\partial \lambda_B^*}{\partial S}|_{S=rk} = 0$ and $\frac{\partial n^*}{\partial S}|_{S=rk} > 0$, we have

$$\begin{aligned}\frac{\partial W}{\partial S}|_{S=rk} &= \frac{1}{r}\frac{\partial n^*}{\partial S}|_{S=rk}\left[\lambda_A^*\frac{\Delta_A}{r} - C_A(\lambda_A^*) + \lambda_B^*\frac{E\Delta_B}{r} - C_B(\lambda_B^*) - rk\right] \\ &> \frac{1}{r}\frac{\partial n^*}{\partial S}|_{S=rk}\left[\lambda_A^* C_A'(\lambda_A^*) - C_A(\lambda_A^*) + \lambda_B^* C_B'(\lambda_B^*) - C_B(\lambda_B^*) - rk\right] \\ &= 0,\end{aligned}$$

where the equality follows from (11).

Summarizing the above discussion, we have:

Remark 2 *As a second-best, the patentability standard that maximizes* $W \equiv W(S)$, S^*, *generally does not maximize the number of firms in the industry. Furthermore, if* $W(S)$ *is single-peaked, then* $S^* > rk$.

Therefore, even though the expected quality improvement from an innovation can be higher in the risky direction than in the safe direction, under the single-peak condition, the welfare-maximizing S does not maximize innovation in the risky direction. This is because by raising S above rk, industry innovation can be increased.

Notice that for S^* to be a valid solution to the maximization problem for $W(S)$, we have implicitly assumed that $S^* \leq \Delta_A$. If this constraint is binding, then we would have $S^* = \Delta_A$. This is because if $S > \Delta_A$, then no entrant would conduct R&D in the safe direction so that $\lambda_A = 0$ and the problem is the same as if the risky direction were the only research direction. But since $C'_A(0) = 0$ by assumption, it is socially desirable to have strictly positive R&D investment in the safe direction. This implies that $W(S)$ would jump down at $S = \Delta_A$. Therefore, it is likely that $S^* \leq \Delta_A$ even if we allow S to be larger than Δ_A.

4.2 Comparing the Innovation Directions

We now compare the equilibrium innovation direction $d(S)$ with the innovation direction that maximizes social welfare, d^o. From (19), we have

$$d^o = \frac{\lambda_B^o}{\lambda_A^o}, \text{ where } \frac{C'(\lambda_B^o)}{C'(\lambda_A^o)} = \frac{E[\Delta_B]}{\Delta_A}.$$

Hence, at the welfare-maximizing innovation direction, the ratio of the marginal costs equals the ratio of the marginal benefits of innovations in the two directions.

Recall from Proposition 1 that the equilibrium innovation direction $d(S)$ is maximized at $S = rk$. The result below states that firms are biased towards (against) innovation in the risky direction when S is below (above) some threshold. A sketch of the proof is as follows:

If $S = 0$, we have $d(S) > d^o$, with a bias towards direction B. As S increases but is smaller than rk, innovation is even more biased towards direction B because, from Proposition 1, λ_B^*/λ_A^* increases in S if $S < rk$. As S further increases and surpasses rk, λ_B^* starts to decrease and λ_A^* to increase, and thus $d(S)$ becomes smaller but can still be larger than d^o. When $S > \hat{S}$, the threshold value of S, $d(S)$ falls below d^o and monotonically decreases, so that innovation direction is biased towards direction A. Formally:

Proposition 4 *There exists $\hat{S} \in \left[rk, \overline{\Delta}_B\right]$ such that $d\left(\hat{S}\right) = d^o$, with $d(S) > d^o$ if $S < \hat{S}$ but $d(S) < d^o$ if $S > \hat{S}$.*

Proof. From (19), under the social optimum,

$$\frac{C_B'\left(\lambda_B^o\right)}{C_A'\left(\lambda_A^o\right)} = \frac{E\Delta_B}{\Delta_A}.$$

From (9) and (10), given S, in the free-entry equilibrium

$$\frac{C_B'\left(\lambda_B^*(S)\right)}{C_A'\left(\lambda_A^*(S)\right)} = \frac{\theta(S)\left[\pi_B(S) - rk\right]}{\Delta_A - rk}.$$

Thus,

$$\delta(S) \equiv \frac{\frac{C_B'\left(\lambda_B^*(S)\right)}{C_A'\left(\lambda_A^*(S)\right)}}{\frac{C_B'\left(\lambda_B^o\right)}{C_A'\left(\lambda_A^o\right)}} = \frac{\theta(S)\left[\pi_B(S) - rk\right]}{E\left[\Delta_B\right]\frac{\Delta_A - rk}{\Delta_A}}.$$

As S increases,

$$\Phi(S) \equiv \theta(S)\left[\pi_B(S) - rk\right]$$

first increases and then decreases, reaching its maximum at $S = rk$, because

$$\begin{aligned}
\frac{\partial\Phi(S)}{\partial S} &= \frac{\partial\theta(S)}{\partial S}\left[\pi_B(S) - rk\right] + \theta(S)\frac{\partial\pi_B(S)}{\partial S} \\
&= -g(S)\left[\pi_B(S) - rk\right] + \theta(S)\frac{g(S)\left[\pi_B(S) - S\right]}{\theta(S)} \\
&= -g(S)(S - rk).
\end{aligned}$$

Moreover, $\Phi(0) = E[\Delta_B] - rk > 0$ and $\Phi\left(\overline{\Delta}_B\right) = 0$. Therefore, there exists a unique $\hat{S} > rk$, determined by

$$\Phi\left(\hat{S}\right) = E[\Delta_B]\frac{\Delta_A - rk}{\Delta_A} < E[\Delta_B],$$

such that $\delta\left(\hat{S}\right) = 1$, with

$$d\left(\hat{S}\right) = \frac{\lambda_B^*\left(\hat{S}\right)}{\lambda_A^*\left(\hat{S}\right)} = \frac{\lambda_B^o}{\lambda_A^o} = d^o.$$

Moreover, since

$$\delta(0) = \frac{\Phi(0)}{E[\Delta_B]\frac{\Delta_A - rk}{\Delta_A}} = \frac{E[\Delta_B] - rk}{E[\Delta_B]\frac{\Delta_A - rk}{\Delta_A}} = \frac{1 - \frac{rk}{E[\Delta_B]}}{1 - \frac{rk}{\Delta_A}} > 1$$

and $\frac{\lambda_B^*(S)}{\lambda_A^*(S)}$—hence $\delta(S)$—increases for $S \in [0, rk)$ but decreases for $S \in \left(rk, \hat{S}\right)$, we have $\delta(S) > 1$ and $\frac{\lambda_B^*(S)}{\lambda_A^*(S)} > d^o$ if $S < \hat{S}$. Also, since $\lambda_B^*(S)$ decreases in S while $\lambda_A^*(S)$ increases in S for $S > rk$, we have $\delta(S) < 1$ and $\frac{\lambda_B^*(S)}{\lambda_A^*(S)} < d^o$ if $S > \hat{S}$. ■

Therefore, $\hat{S}$ implements the welfare-maximizing innovation direction, provided that $\hat{S} < \Delta_A$; innovation is biased towards B when $S < \hat{S}$, whereas it is biased towards A when $S > \hat{S}$. To understand this result, note that the innovation direction that maximizes social welfare, d^o, is invariant with patentability standard because all (patentable and unpatentable) innovations will increase social benefits. However, market incentives do change as patentability standard varies. Specifically, when the patentability standard is relatively low, the risky research direction with uncertain innovation size is likely to yield a patent even when the quality improvement is small, increasing the innovation incentive through that direction. Reinforced by the incumbency-prolonging and strategic substitution effect, this motivates firms to conduct R&D in the risky direction excessively relative to the direction with a certain innovation size. Conversely, when the patentability standard is high enough, the direction with uncertain innovation size is unlikely to receive a patent even when the quality improvement is relatively large, which unduly discourages R&D in that direction.

Notice that while $\hat{S}$ leads to the efficient choice of research direction, it need not be the welfare-maximizing S for $W(S)$ in the second-best problem. This is because S also affects $W(S)$ through $n^*(S)$, as can be seen from (22), and thus $\hat{S}$ need not maximize $W(S)$. Intuitively, the second-best choice of S, S^*, will generally involve a trade off between two policy goals: moving towards the efficient R&D direction (d^o) and towards the efficient number of entrants (n^o). When S achieves the efficient R&D direction, as $\hat{S}$ does, it does not optimally balance the two goals, and hence in general $\hat{S}$ does not maximize $W(S)$ (i.e. $\hat{S} \neq S^*$).

5 Discussion

For tractability, our model has made the restrictive assumptions that the incumbent does not conduct R&D and that the patentability standard is exogenously given and fixed. In this section, we discuss how our analysis might change if these assumptions are relaxed.

In subsection 5.1, we illustrate, in a particular setting and with an example, that the main insights of our model can still be valid when incumbents also conduct R&D. Subsection 5.2 further illustrates how our results might change if patentability standards are state-dependent.

5.1 When Incumbents Also Conduct R&D

We have shown how patentability standards impact innovation through the threshold, incumbency-prolonging, strategic substitution and market-structure effects in a model where each incumbent (the leader) is assumed not to innovate. The analysis for the situation where the incumbent also conducts R&D is generally very complicated. This is because the incumbent may invest different amounts of R&D after its successful innovation each time, and if the incumbent keeps (luckily) succeeding there are infinitely many states and R&D strategies for the incumbent. In this case, there is no stationary equilibrium.

To get a flavor of the analysis when incumbents can innovate, suppose that each incumbent can conduct R&D in all periods of its incumbency and profit from all of its patented technologies in a cumulative sense, until it is replaced by a challenger whose R&D starts from the current state of technology. Then, an infinite number of equations are needed to describe the dynamic system in equilibrium, starting from

$$rV^E = \lambda_A \left(V_A^I - V^E\right) - C_A(\lambda_A) + \theta\lambda_B \left(V_B^I - V^E\right) - C_B(\lambda_B), \tag{23}$$

where V_z^I, $z \in \{A, B\}$ is, as before, the expected value of being an incumbent through type-z innovation. These expected values depend on the incumbent's further innovation. Hence:

$$\begin{aligned} rV_A^I &= \Delta_A + n(\lambda_A + \theta\lambda_B)\left(V^E - V_A^I\right) \\ &\quad + \lambda_{AA}\left(V_{AA}^I - V_A^I\right) - C_A(\lambda_{AA}) + \theta\lambda_{BA}\left(V_{BA}^I - V_A^I\right) - C_B(\lambda_{BA}), \end{aligned} \tag{24}$$

and

$$\begin{aligned} rV_B^I &= \pi_B + n(\lambda_A + \theta\lambda_B)\left(V^E - V_B^I\right) \\ &\quad + \lambda_{AB}\left(V_{AB}^I - V_B^I\right) - C_A(\lambda_{AB}) + \theta\lambda_{BB}\left(V_{BB}^I - V_B^I\right) - C_B(\lambda_{BB}), \end{aligned} \tag{25}$$

where $V_{z_1 z}^I$ denotes the expected value of maintaining the leader position through type-z_1 innovation for the incumbent who currently has a type-z technology, with z_1 and $z \in \{A, B\}$.

But the values of $V_{z_1 z}^I$ depend on the incumbent's further innovation, as described by (AC-1) in Appendix C, where $V_{z_2 z_1 z}^I$ is the expected value of maintaining the leader position through type-z_2 innovation for the incumbent who currently has a type-z_1 and a type-z technology, for z_2, z_1, and $z \in \{A, B\}$. This process can potentially continue indefinitely.

To gain insights on how the effects might change when incumbents are allowed to innovate in a setting that is still tractable, we next consider the case where the incumbent conducts R&D until it succeeds at most one more time (i.e. it can succeed at most two consecutive periods before stopping R&D). Then, all values of $V_{z_2 z_1 z}^I$ in (AC-1) are zero. In this case, the quality size by which the incumbent leads is $\Delta_z + \Delta_{z_1}$.

From (23)–(25), we can obtain the first-order conditions on λ_A, λ_B, λ_{AA}, λ_{AB}, λ_{BA}, and λ_{BB}, as in (AC-2) of Appendix C. Substituting these conditions back into (23)–(25) and (AC-1), and using the free-entry condition $V^E = k$, we obtain a system of seven equations, as in (AC-3) of Appendix C, which determines the seven equilibrium values λ_A^*, λ_B^*, λ_{AA}^*, λ_{BA}^*, λ_{AB}^*, λ_{BB}^* and n^*. Appendix C also contains the expressions for industry innovation direction $d(S)$, industry innovation rate R, and total welfare W.

While we are unable to obtain analytical solutions to the system of equations in (AC-3), we can numerically compute the equilibrium. Figure 2 plots $\lambda_A^*(S)$, $\lambda_B^*(S)$, $d^*(S)$, $n^*(S)$, $R^*(S)$, and $W^*(S)$ for $S \in [0, 0.3)$ of Example 2 in Appendix C. The curves in Figure 2 have patterns that are broadly consistent with those in Figure 1, especially at the industry level. Thus, our main results on how patentability standards impact innovation can be valid when incumbents are allowed to engage in R&D as well. This is not entirely surprising because, intuitively, the incumbent itself can be viewed as a new challenger, whose participation in the patent race need not change the basic forces at work.

One notable difference in Figure 2 is that as S increases, $\lambda_A^*(S)$ barely decreases initially before increasing, and $\lambda_B^*(S)$ barely increases initially before decreasing, while the other key relationships appear to have much more similar shapes as those in Figure 1. Thus, allowing incumbents to also conduct R&D appears to make (initial) increases in S more negatively impact innovation in the risky direction and more positively impact innovation in the safe direction. Recall that when incumbents are assumed to conduct no R&D, the incumbency-prolonging effect plays a key role in leading to the initial decrease of $\lambda_A^*(S)$ and initial increase of $\lambda_B^*(S)$. When incumbents can also innovate, a reduction of R&D by new entrants has less impact on the duration of the incumbency, which weakens the incumbency-prolonging effect. This, together with the interaction with the threshold and strategic substitution effects, appears to be responsible for the main difference between the two figures.[22]

5.2 State-Dependent Patentability Standards

Our model has assumed that the patentability standard is exogenously given and fixed. It would also be interesting to examine the situation where the patentability standard—the quality improvement required to be granted a patent—is endogenously determined and depends on the current technology state of the industry.[23] One major difficulty, however, is that firms' R&D strategies will no longer be stationary under state-dependent patentability standards, which makes it a formidable task to analyze the problem generally.

To illustrate some possible new insights and implications, we consider a particular state-dependent patentability policy under which the current patentability standard is set as high as the one in the patent that was previously granted (Amano, 2016). Specifically, the state-dependent patentability policy defines a patentability standard in the risky direction, S_t, at time t as

$$S_t = D_B\left(t\right),$$

where $D_B\left(t\right)$ is the size of the most recent innovation (at time t) that has been granted a patent. Note that in our main model with fixed patentability, S_t is invariant over time.

Our first—and obvious—observation is that under the state-dependent patentability policy, the patentability standard will gradually increase over time.

We next compare innovation incentives under the fixed and state-dependent patentability policies. In particular, we consider the question: at any moment t, if innovating firms face the same patentability standard under the two policies, which policy will induce higher R&D

[22] Our specific formulation of the incumbent innovation assumes that an incumbent increases its quality lead additively when it has multiple innovations. Suppose instead that when an incumbent succeeds in a new innovation, its earlier lead is partially replaced, possibly due to leakages of its earlier technology. Then, the usual replacement effect will become more pronounced, reducing the incentives of incumbent innovation. Consequently, the incumbency-prolonging effect—and our results— would be less affected if incumbents are allowed to conduct R&D.

[23] Acemoglu and Akcigit (2012) take into account state-dependent patent length in a framework of step-by-step innovation.

investment?

As in our main model, the innovation incentives here are mainly determined by the threshold, incumbency-prolonging, and strategic substitution effects. Under the same patentability standard, a firm faces the same difficulty in obtaining a patent from innovation (the threshold effect). However, the incumbency-prolonging effect is stronger under the state-dependent patentability policy because an innovating firm expects a higher future patentability standard after its successful patentable innovation and thus a longer incumbency. Hence, the R&D intensity is higher in the risky direction under state-dependent patentability policy due to a stronger incumbency-prolonging effect. Moreover, the R&D incentive is lower in the safe direction because a higher R&D intensity in one direction reduces the innovation incentive of the other direction (the strategic substitution effect). We thus have the following remark.

Remark 3 *Given a patentability standard S_t, firms have higher (lower) incentives to invest in the risky (safe) direction under the state-dependent patentability policy than under the fixed patentability policy.*

We may also ask how equilibrium R&D intensities in different directions and industry innovation change with the endogenously determined patent policy over time. Recall that in our model with a fixed patentability standard, R&D intensities in the risky and safe directions are time-invariant but—from Proposition 1—as S rises, R&D intensity in the risky (safe) direction first increases (decreases) and then decreases (increases). Moreover, from Proposition 2, as S rises, innovation innovation initially increases and eventually decreases. Since the patentability standard increases over time under the state-dependent patentability policy, Propositions 1 and 2 imply the following:

Remark 4 *Under the state-dependent patentability policy, it is possible that, over time, R&D intensity in the risky (safe) direction first increases (decreases) and eventually decreases (increases). Moreover, industry innovation will initially increase but eventually de-*

crease.

6 CONCLUSION

This paper has provided a first look at how patent policy may impact the rate and direction of cumulative innovation when firms can conduct R&D in multiple directions. We have three main findings: (i) Patentability standards affect the rate of industry innovation through both the number of entrants and their R&D intensities in the free entry equilibrium. As S rises, the rate of industry innovation initially increases and eventually decreases. (ii) Compared to the social optimum, market incentives for cumulative innovation are deficient for both R&D intensities and the number of entrants. (iii) There exists a critical level of patentability standard ($\hat{S}$) under which the innovation direction is efficient, whereas R&D is biased towards (against) the risky direction when S is below (above) $\hat{S}$. However, if S is the only policy variable available, then the optimal policy, which balances the trade-off between the rate and direction of innovation, will in general be different from $\hat{S}$.

Discussions about patent policy and the patent system have frequently surrounded the issue of patentability standards. It has been argued that patentability standards in the U.S. are too low, leading to excessive incentives for small-size innovations (e.g., Hunt, 2004; Jaffe, 2000). Our results suggest that raising patentability standards may indeed improve innovation direction, with two caveats: first, the effect of a higher S on innovation direction may be non-monotonic, and a small increase in S can either alleviate or exacerbate possible direction biases depending on the starting point; second, in our model, the risky direction may lead to more small-size innovations but to a higher expected size than the safe direction. Hence, even when raising S reduces the patenting of small-size innovations, it may not raise the expected innovation size.

In our model, the fixed setup cost for R&D (adjusted by r), rk, plays important roles in determining the innovation incentives and the optimal patentability standards. This cost generally differs for different industries. For instance, it is likely much larger in the pharmaceutical industry than in the software industry. Thus, it would be desirable that patentability standards differ for different industries, depending (indirectly) on the setup cost for R&D projects. Moreover, innovations in developing countries tend to be much below the world technology frontier and require lower setup cost rk than those in developed countries. Then, the desirable patentability requirement could be lower in developing countries in order to promote innovation.[24]

[24] Chen and Puttitanun (2005) shows how intellectual property rights (IPRs) affect innovations in developing countries and how the optimal IPRs policy may vary with a country's level of development.

References

[1] Acemoglu, D. (2002). “Directed Technical Change.”, *Review of Economic Studies*, 69, 781-809.

[2] Acemoglu, D. (2011). “Diversity and Technological Progress” in "The Rate and Direction of Inventive Activity Revisited", edited by Scott Stern and Josh Lerner, 319-356. University of Chicago Press.

[3] Acemoglu, D. and U. Akcigit. (2012). “Intellectual Property Rights Policy, Competition and Innovation”, *Journal of the European Economic Association*, 10, 1-42.

[4] Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt. (2005). “Competition and Innovation: An Inverted-U Relationship”, *Quarterly Journal of Economics*, 120, 701-728.

[5] Aghion, P., D. Dewatripont and J. Stein. (2008). “Academic Freedom, Private-Sector Focus, and the Process of Innovation”, *RAND Journal of Economics*, 39, 617-635.

[6] Amano, T., (2016), “Ratcheting, Competition, and the Diffusion of Technological Change: The Case of Televisions under an Energy Efficiency Program”, Working paper.

[7] Anderson, Axel, and Luís Cabral. (2007). “Go for Broke or Play it Safe? Dynamic Competition with Choice of Variance”, *Rand Journal of Economics*, 38, 593-609.

[8] Bessen, J. and E. Maskin. (2009). “Sequential Innovation, Patents, and Imitation”, *RAND Journal of Economics*, 31, 611-635.

[9] Bhattacharya, S. and D. Mookherjee. (1986). “Portfolio Choice in Research and Development”, *RAND Journal of Economics*, 17, 594-605.

[10] Birkinshaw, J. and M. Lingblad. (2001). “Making Sense of Internal Competition: Designs for Organizational Redundancy in Response to Environmental Uncertainty”, London Business School Working Paper.

[11] Bryan, K. and J. Lemus. (2016). "The Direction of Innovation", Working paper.

[12] Cabral, L. (1994). "Bias in Market R&D Portfolios", *International Journal of Industrial Organization*, 12, 533-547.

[13] Chen, Y., S. Pan and T. Zhang. (2014). "(When) Do Stronger Patents Stimulate Continual Innovation?", *Journal of Economic Behavior & Organization*, 98, 115-124.

[14] Chen, Y. and T. Puttitanun. (2005). "Intellectual Property Rights and Innovation in Developing Countries", *Journal of Development Economics*, 78, 474-493.

[15] Chiang, A. and K. Wainwright (2005). *Fundamental Methods of Mathematical Economics*. McGraw-Hill.

[16] Choi, J. and H. Gerlach. (2014). "Selection Biases in Complementary R&D Projects", *Journal of Economics & Management Strategy*, 23, 899-924.

[17] Dasgupta, P. and E. Maskin. (1987). "The Simple Economics of Research Portfolio", *Economic Journal*, 97, 581-595.

[18] Furman, Jeffrey and Scott Stern. (2011). "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research", *American Economic Review*, 101, 1933-1963.

[19] Galasso, Alberto and Mark Schankerman. (2013). "Patents and Cumulative Innovation: Causal Evidence from the Courts", Working Paper.

[20] Hopenhayn, H. and F. Squintani. (2016). "The Direction of Innovation", Working paper.

[21] Horowitz, A. and E. Lai. (1996). "Patent Length and the Rate of Innovation", *International Economic Review*, 37, 785-801.

[22] Hunt, Robert M., (2004). "Patentability, Industry Structure, and Innovation", *Journal of Industrial Economics*, 52, 401-425.

[23] Jaffe, A., (2000). "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process", *Research Policy*, 29, 531-557.

[24] Klette, T. and D. de Meza. (1986). "Is the Market Biased against Risky R&D?", *Rand Journal of Economics*, 17, 133-139.

[25] Kwon, I. (2010). "R&D Portfolio and Market Structure", *Economic Journal*, 120, 313-323.

[26] Lee, T. and L. Wilde. (1980). "Market and Structure and Innovation: A Rewformulation", *Quarterly Journal of Economics*, 94, 429-436.

[27] Loury, G., (1979). "Market Sturcture and Innovation", *Quarterly Journal of Economics*, 93, 395-410.

[28] Murray, F. and S. Stern. (2007). "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-commons Hypothesis", *Journal of Economic Behavior and Organization*, 63, 648-687.

[29] O'Donoghue, Ted. (1998). "A Patentability Requirement for Sequential Innovation", *RAND Journal of Economics*, 29, 654-679.

[30] O'Donoghue, Ted, Suzanne Scotchmer, and Jacques-Francois Thisse. (1998). "Patent Breadth, Patent Life, and the Pace of Technological Progress", *Journal of Economics and Management Strategy*, 7, 1-32.

[31] Sampat, Bhaven and H. Williams. (2014). "How Do Patents Affect Follow-on Innovation? Evidence from the Human Genome", Working Paper.

[32] Segal, I. and M. Whinston. (2007). "Antitrust in Innovative Industries", *American Economic Review*, 97, 1703-1730.

[33] Williams, Heidi. (2013). "Intellectual Property Rights and Innovation: Evidence from the Human Genome", *Journal of Political Economy*, 121, 1-27.

Appendix A: An Illustrative Example

We illustrate the main theoretical results in this paper through Example 1 below.

Example 1. Suppose that Δ_B follows the uniform distribution on $[0,1]$, while $C_A(\lambda_A) = \frac{1}{2}\lambda_A^2$ and $C_B(\lambda_B) = 5\lambda_B^2$. Then, from (9), (10) and (11):

$$\lambda_A^* = \sqrt{\frac{2rk}{1 + 0.1\left[\frac{(1-S)(\frac{1+S}{2}-rk)}{\Delta_A - rk}\right]^2}},$$

$$\lambda_B^* = \sqrt{\frac{2rk}{10 + 100\left[\frac{\Delta_A - rk}{(1-S)(\frac{1+S}{2}-rk)}\right]^2}},$$

and

$$n^* = \left(\frac{\Delta_A - rk}{\lambda_A^*} - r\right) \cdot \frac{1}{\lambda_A^* + (1-S)\lambda_B^*}.$$

Assume $r = 0.08$, $k = 1.2$, $\Delta_A = 0.3$, and let $S = 0.1$ which is close to $rk = 0.096$. Then $\lambda_A^* = 0.3701$, $\lambda_B^* = 0.0742$, and $n^* = 1.0785$.

(a). First, S is allowed to vary. Figure 1 below plots λ_A^*, λ_B^*, d, n^*, R and W as functions of $S \in [0, 0.3)$, where $\lambda_A^*(S)$, $\lambda_B^*(S)$ and $d(S)$ reach the minimum 0.3701 and the maxima 0.0742 and 0.2004, respectively, at $S_{\{\lambda_A^*,\lambda_B^*,d\}} = rk = 0.096$, illustrating Proposition 1.

(b). Next, $R(S)$ in Figure 1 reaches the maxima 0.1618 at $S_R = 0.2494$, greater than $rk = 0.096$, illustrating Proposition 2 and Remark 1.

(c). In Figure 1, $\lambda_A^*(S)$, $\lambda_B^*(S)$ and $n^*(S)$ reach the maxima 0.3755, 0.0742 and 1.0912 at $S_{max} = 0.3$, $S_{\{\lambda_A^*,\lambda_B^*,d\}} = 0.096$ and $S_{n^*} = 0.24$, respectively. Therefore, $\lambda_A^*(S) < \lambda_A^o = \frac{\Delta_A}{r} = 3.75$, $\lambda_B^*(S) < \lambda_B^o = \frac{E(\Delta_B)}{r} = 0.625$, and $n^*(S) < n^o = \infty$, for all $S \in [0, 0.3)$. This illustrates Proposition 3 for all feasible $S < \Delta_A = 0.3$. That is, compared to the first-best, R&D intensities and the number of entrants under the free entry equilibrium are deficient, no mater where the feasible patentability standard is.

(d). Moreover, $W(S)$ in Figure 1 reaches the maximum 22.6586 at $S^* = 0.2506$, which is not equal to $S_{n^*} = 0.24$, and is greater than $rk = 0.096$. Then, as a second-best, the patentability standard that maximizes $W(S)$ does not maximize the number of firms in the industry, and is greater than the patentability standard that minimizes $\lambda_A^*(S)$ and maximizes $\lambda_B^*(S)$ and $d(S)$. This illustrates Remarks 2.

(e). Notice that

$$\theta(\pi_B - rk) = (1-S)\left(\frac{\int_S^1 x dx}{1-S} - 0.096\right) = (1-S)\left(\frac{1+S}{2} - 0.096\right),$$

and

$$E[\Delta_B]\frac{\Delta_A - rk}{\Delta_A} = 0.34.$$

Thus, from

$$(1-S)\left(\frac{1+S}{2} - 0.096\right) = 0.34,$$

we obtain $\hat{S}^e = 0.4664$, greater than $rk = 0.096$ and Δ_A, such that $\hat{S}^e$ is a corner solution to Proposition 4, as $S \in [0, 0.3)$ in this example.

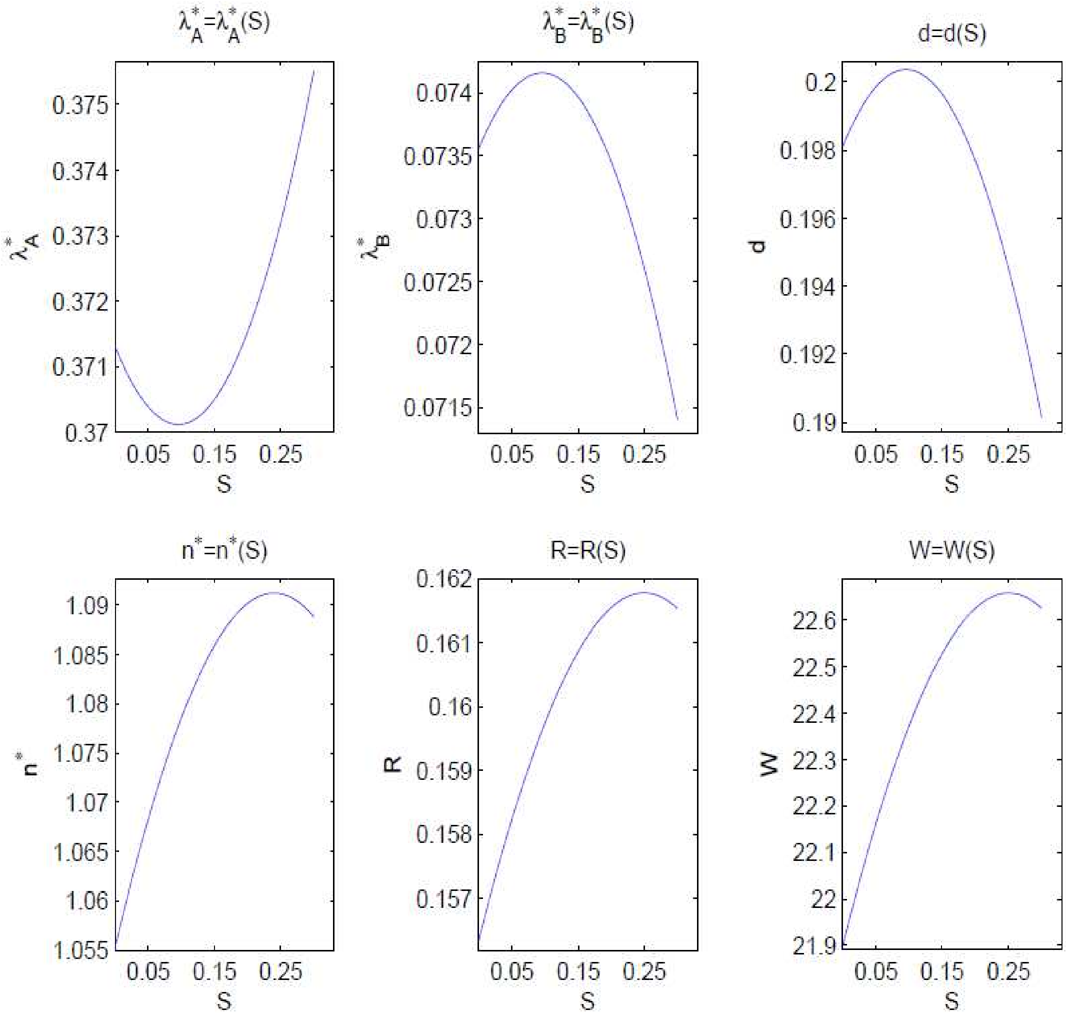

Figure 1: Example 1

Appendix B: Proofs

Appendix B contains proofs for Proposition 1, Proposition 2, and (17).

Proof of Proposition 1. From (3), (4) and (5), the equilibrium λ_A^*, λ_B^* and n^* solve the system of equations below:

$$M \equiv \begin{bmatrix} M^1 \equiv C_A'(\lambda_A)\left[r + n(\lambda_A + \theta\lambda_B)\right] - (\pi_A - rk) \\ M^2 \equiv \lambda_A C_A'(\lambda_A) + \lambda_B C_B'(\lambda_B) - C_A(\lambda_A) - C_B(\lambda_B) - rk \\ M^3 \equiv \theta C_A'(\lambda_A)(\Pi_B - rk) - C_B'(\lambda_B)(\pi_A - rk) \end{bmatrix} = 0$$

Let $M_j^i \equiv \frac{\partial M^i}{\partial j}$, for $i = 1, 2, 3$ and $j = \lambda_A, \lambda_B, S, n$. Define $|M_{AS}|, |M_{BS}|, |M_{nS}|$ and $|M|$ as the determinants of matrix M_{AS}, M_{BS}, M_{nS} and M, respectively:

$$|M_{AS}| = \begin{vmatrix} M_S^1 & M_{\lambda_B}^1 & M_n^1 \\ M_S^2 & M_{\lambda_B}^2 & M_n^2 \\ M_S^3 & M_{\lambda_B}^3 & M_n^3 \end{vmatrix}, \qquad |M_{BS}| = \begin{vmatrix} M_{\lambda_A}^1 & M_S^1 & M_n^1 \\ M_{\lambda_A}^2 & M_S^2 & M_n^2 \\ M_{\lambda_A}^3 & M_S^3 & M_n^3 \end{vmatrix},$$

$$|M_{nS}| = \begin{vmatrix} M_{\lambda_A}^1 & M_{\lambda_B}^1 & M_S^1 \\ M_{\lambda_A}^2 & M_{\lambda_B}^2 & M_S^2 \\ M_{\lambda_A}^3 & M_{\lambda_B}^3 & M_S^3 \end{vmatrix} \quad \text{and} \quad |M| = \begin{vmatrix} M_{\lambda_A}^1 & M_{\lambda_B}^1 & M_n^1 \\ M_{\lambda_A}^2 & M_{\lambda_B}^2 & M_n^2 \\ M_{\lambda_A}^3 & M_{\lambda_B}^3 & M_n^3 \end{vmatrix}.$$

By Cramer's rule[25], we have

$$\frac{\partial \lambda_A^*}{\partial S} = -\frac{|M_{AS}|}{|M|} \qquad \text{and} \qquad \frac{\partial \lambda_B^*}{\partial S} = -\frac{|M_{BS}|}{|M|}.$$

[25] Cramer's rule is a standard way of finding the partial derivative of an independent variable in a system of equations. See, for example, Chiang and Wainwright (2005) for an introduction to Cramer's rule.

We next compute the relevant derivatives:

$$
\begin{aligned}
M^1_{\lambda_A} &= [r + n(\lambda_A + \theta\lambda_B)]C''_A(\lambda_A) + nC'_A(\lambda_A), \quad M^1_{\lambda_B} = n\theta C'_A(\lambda_A), \\
M^1_n &= (\lambda_A + \theta\lambda_B)C'_A(\lambda_A), \quad M^1_S = -g(S)n\lambda_B C'_A(\lambda_A), \\
M^2_{\lambda_A} &= \lambda_A C''_A(\lambda_A), \quad M^2_{\lambda_B} = \lambda_B C''_B(\lambda_B), \\
M^2_n &= 0, \quad M^2_S = 0, \quad M^3_{\lambda_A} = \theta(\Pi_B - rk)C''_A(\lambda_A), \quad M^3_{\lambda_B} = -(\pi_A - rk)C''_B(\lambda_B), \\
M^3_n &= 0, \quad M^3_S = g(S)C'_A(\lambda_A)(rk - S).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
|M_{AS}| &= M^1_S M^2_{\lambda_B} M^3_n + M^1_{\lambda_B} M^2_n M^3_S + M^1_n M^2_S M^3_{\lambda_B} - M^1_n M^2_{\lambda_B} M^3_S - M^1_{\lambda_B} M^2_S M^3_n - M^1_S M^2_n M^3_{\lambda_B} \\
&= -g(S)\lambda_B(\lambda_A + \theta\lambda_B)\left(C'_A\right)^2 C''_B(rk - S), \\
|M_{BS}| &= M^1_{\lambda_A} M^2_S M^3_n + M^1_S M^2_n M^3_{\lambda_A} + M^1_n M^2_{\lambda_A} M^3_S - M^1_n M^2_S M^3_{\lambda_A} - M^1_S M^2_{\lambda_A} M^3_n - M^1_{\lambda_A} M^2_n M^3_S \\
&= g(S)\lambda_A(\lambda_A + \theta\lambda_B)\left(C'_A\right)^2 C''_A(rk - S),
\end{aligned}
$$

and

$$
\begin{aligned}
|M| &= M^1_{\lambda_A} M^2_{\lambda_B} M^3_n + M^1_{\lambda_B} M^2_n M^3_{\lambda_A} + M^1_n M^2_{\lambda_A} M^3_{\lambda_B} - M^1_n M^2_{\lambda_B} M^3_{\lambda_A} - M^1_{\lambda_B} M^2_{\lambda_A} M^3_n - M^1_{\lambda_A} M^2_n M^3_{\lambda_B} \\
&= -(\lambda_A + \theta\lambda_B)C'_B C''_A C''_B[\lambda_A(\pi_A - rk) + \theta\lambda_B(\Pi_B - rk)] < 0.
\end{aligned}
$$

Therefore,

$$\frac{\partial\lambda^*_A}{\partial S} = -\frac{|M_{AS}|}{|M|} = \frac{g(S)\lambda^*_B(\lambda^*_A + \theta\lambda^*_B)\left(C'_A\right)^2 C''_B(rk - S)}{|M|},$$

and

$$\frac{\partial\lambda^*_B}{\partial S} = -\frac{|M_{BS}|}{|M|} = \frac{-g(S)\lambda^*_A(\lambda^*_A + \theta\lambda^*_B)\left(C'_A\right)^2 C''_A(rk - S)}{|M|}.$$

It follows that $\frac{\partial\lambda^*_B}{\partial S} > 0$ and $\frac{\partial\lambda^*_A}{\partial S} < 0$ if $S < rk$ and $\frac{\partial\lambda^*_B}{\partial S} < 0$ and $\frac{\partial\lambda^*_A}{\partial S} > 0$ if $rk < S$. This further implies that λ^*_B/λ^*_A, same as λ^*_B, is an inverted-U function of S, maximized at

$S = rk$. ■

Proof of Proposition 2. By Cramer's rule,

$$\frac{\partial n}{\partial S} = -\frac{|M_{nS}|}{|M|}.$$

where

$$\begin{aligned} |M_{nS}| &= g(S)C_A' C_A'' C_B'' \left\{ \begin{array}{c} n\lambda_B[\lambda_A(\pi_A - rk) + \theta\lambda_B(\Pi_B - rk)] \\ +\lambda_B(rk - S)[r + n(\lambda_A + \theta\lambda_B)] \end{array} \right\} \\ &\quad + g(S)n\left(C_A'\right)^2 \left(\lambda_B C_B'' - \theta\lambda_A C_A''\right)](rk - S). \end{aligned} \tag{AB-1}$$

From (13), (14) and (AB-1), we can show that, after substitution and simplification,

$$\begin{aligned} \frac{\partial R}{\partial S} &= (\lambda_A\Delta_A + \lambda_B E[\Delta_B])\frac{\partial n}{\partial S} + \Delta_A\frac{\partial \lambda_A}{\partial S} + E[\Delta_B]\frac{\partial \lambda_B}{\partial S} \\ &= \frac{-g(S)C_A' \left\{ \begin{array}{c} (\lambda_A\Delta_A + \lambda_B E[\Delta_B])C_A'' C_B'' \lambda_B[n\lambda_A(\pi_A - S) \\ +n\theta\lambda_B(\Pi_B - S) + r(rk - S)] \\ +nC_A'(E[\Delta_B] - \theta\Delta_A)\left(\lambda_A^2 C_A'' + \lambda_B^2 C_B''\right)(rk - S) \end{array} \right\}}{|M|}. \end{aligned} \tag{AB-2}$$

Note that $E[\Delta_B] \geq \theta\Delta_A$ since $\theta \leq 1$. If $S = 0$, then $S < \min\{\pi_A, \Pi_B, rk\}$. It follows that $\frac{\partial R}{\partial S}|_{S=0} > 0$. If $S = \bar{\Delta}_B$, then $S \geq \max\{\pi_A, \Pi_B, rk\}$ and we have $\frac{\partial R}{\partial S}|_{S=\bar{\Delta}_B} < 0$. ■

Proof of (17). From (AB-2),

$$\left.\frac{\partial R}{\partial S}\right|_{S=rk} = \frac{-g(S)C_A'}{|M|}(\lambda_A\Delta_A + \lambda_B E\Delta_B)C_A'' C_B'' \lambda_B[n\lambda_A(\pi_A - rk) + n\theta\lambda_B(\Pi_B - rk)] > 0.$$

That is (17). ■

Appendix C

$$\begin{cases} rV^I_{AA} = \Delta_A + \Delta_A + n\left(\lambda_A + \theta\lambda_B\right)\left(V^E - V^I_{AA}\right) \\ +\lambda_{AAA}\left(V^I_{AAA} - V^I_{AA}\right) - C_A\left(\lambda_{AAA}\right) + \theta\lambda_{BAA}\left(V^I_{BAA} - V^I_{AA}\right) - C_B\left(\lambda_{BAA}\right), \\ rV^I_{BA} = \pi_B + \Delta_A + n\left(\lambda_A + \theta\lambda_B\right)\left(V^E - V^I_{BA}\right) \\ +\lambda_{ABA}\left(V^I_{ABA} - V^I_{BA}\right) - C_A\left(\lambda_{ABA}\right) + \theta\lambda_{BBA}\left(V^I_{BBA} - V^I_{BA}\right) - C_B\left(\lambda_{BBA}\right), \\ rV^I_{AB} = \Delta_A + \pi_B + n\left(\lambda_A + \theta\lambda_B\right)\left(V^E - V^I_{AB}\right) \\ +\lambda_{AAB}\left(V^I_{AAB} - V^I_{AB}\right) - C_A\left(\lambda_{AAB}\right) + \theta\lambda_{BAB}\left(V^I_{BAB} - V^I_{AB}\right) - C_B\left(\lambda_{BAB}\right), \\ rV^I_{BB} = \pi_B + \pi_B + n\left(\lambda_A + \theta\lambda_B\right)\left(V^E - V^I_{BB}\right) \\ +\lambda_{ABB}\left(V^I_{ABB} - V^I_{BB}\right) - C_A\left(\lambda_{ABB}\right) + \theta\lambda_{BBB}\left(V^I_{BBB} - V^I_{BB}\right) - C_B\left(\lambda_{BBB}\right), \end{cases} \tag{AC-1}$$

$$\begin{cases} V^I_A - V^E = C'_A\left(\lambda_A\right), \quad \theta\left(V^I_B - V^E\right) = C'_B\left(\lambda_B\right), \\ V^I_{AA} - V^I_A = C'_A\left(\lambda_{AA}\right), \quad \theta\left(V^I_{BA} - V^I_A\right) = C'_B\left(\lambda_{BA}\right), \\ V^I_{AB} - V^I_B = C'_A\left(\lambda_{AB}\right), \quad \theta\left(V^I_{BB} - V^I_B\right) = C'_B\left(\lambda_{BB}\right), \end{cases} \tag{AC-2}$$

and

$$\begin{cases} rk = \lambda_A C'_A\left(\lambda_A\right) - C_A\left(\lambda_A\right) + \lambda_B C'_B\left(\lambda_B\right) - C_B\left(\lambda_B\right), \\ \left[r + n(\lambda_A + \theta\lambda_B)\right] C'_A\left(\lambda_A\right) = \Delta_A - rk + \lambda_{AA} C'_A\left(\lambda_{AA}\right) - C_A\left(\lambda_{AA}\right) \\ +\lambda_{BA} C'_B\left(\lambda_{BA}\right) - C_B\left(\lambda_{BA}\right), \\ \left[r + n(\lambda_A + \theta\lambda_B)\right] C'_B\left(\lambda_B\right) = \theta\pi_B - \theta rk + \theta\lambda_{AB} C'_A\left(\lambda_{AB}\right) - \theta C_A\left(\lambda_{AB}\right) \\ +\theta\lambda_{BB} C'_B\left(\lambda_{BB}\right) - \theta C_B\left(\lambda_{BB}\right), \\ \left[r + n(\lambda_A + \theta\lambda_B)\right]\left[C'_A\left(\lambda_{AA}\right) + C'_A\left(\lambda_A\right)\right] = 2\Delta_A - rk, \\ \left[r + n(\lambda_A + \theta\lambda_B)\right]\left[C'_B\left(\lambda_{BA}\right) + \theta C'_A\left(\lambda_A\right)\right] = \theta\pi_B + \theta\Delta_A - \theta rk, \\ \left[r + n(\lambda_A + \theta\lambda_B)\right]\left[\theta C'_A\left(\lambda_{AB}\right) + C'_B\left(\lambda_B\right)\right] = \theta\Delta_A + \theta\pi_B - \theta rk, \\ \left[r + n(\lambda_A + \theta\lambda_B)\right]\left[C'_B\left(\lambda_{BB}\right) + C'_B\left(\lambda_B\right)\right] = 2\theta\pi_B - \theta rk. \end{cases} \tag{AC-3}$$

The overall innovation direction of the industry is

$$d(S) = \frac{\begin{array}{c}\lambda_B^* + \lambda_A^*\lambda_{BA}^* \min\left\{\frac{1}{\lambda_{AA}^*+\theta\lambda_{BA}^*}, \frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}\right\} \\ +\theta\lambda_B^*\lambda_{BB}^* \min\left\{\frac{1}{\lambda_{AB}^*+\theta\lambda_{BB}^*}, \frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}\right\}\end{array}}{\begin{array}{c}\lambda_A^* + \lambda_A^*\lambda_{AA}^* \min\left\{\frac{1}{\lambda_{AA}^*+\theta\lambda_{BA}^*}, \frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}\right\} \\ +\theta\lambda_B^*\lambda_{AB}^* \min\left\{\frac{1}{\lambda_{AB}^*+\theta\lambda_{BB}^*}, \frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}\right\}\end{array}}, \tag{AC-4}$$

where $\frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}$ is the expected incumbency, $n^*\lambda_z^*$ is the R&D intensity in direction $z \in \{A, B\}$ for the n^* challengers per period, $\frac{\lambda_A^*}{\lambda_A^*+\theta\lambda_B^*}$ is the probability that an incumbent replaces the previous incumbent through type-A innovation, $\frac{\theta\lambda_B^*}{\lambda_A^*+\theta\lambda_B^*}$ is the probability that an incumbent replaces the previous incumbent through type-B innovation, $\min\left\{\frac{1}{\lambda_{AA}^*+\theta\lambda_{BA}^*}, \frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}\right\}$ is the effective time for the incumbent who currently holds a type-A patented technology to conduct R&D until it successfully innovates a patented technology or is replaced by a challenger[26], $\min\left\{\frac{1}{\lambda_{AB}^*+\theta\lambda_{BB}^*}, \frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}\right\}$ is the effective time for the incumbent who currently holds a type-B patented technology to conduct R&D until it successfully innovates a patented technology or is replaced by a challenger, and $\lambda_{z'z}^*$ is the R&D intensity in direction $z' \in \{A, B\}$ for the incumbent who currently holds a type-z patented technology with $z \in \{A, B\}$.

Similar to the above reasoning, the overall innovation rate of the industry, R, and total welfare, W, can be respectively determined as follows:

[26] If the incumbent currently holds a type-A patented technology, $\frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}$ is the expected incumbency, and $\frac{1}{\lambda_{AA}^*+\theta\lambda_{BA}^*}$ is the expected duration for the incumbent to conduct R&D until it successfully innovates a patented technology. The incumbent will conduct R&D until it is replaced by a challenger or successfully innovates a patented technology. So the effective time for the incumbent to conduct R&D is the smaller of $\frac{1}{n^*(\lambda_A^*+\theta\lambda_B^*)}$ and $\frac{1}{\lambda_{AA}^*+\theta\lambda_{BA}^*}$.

$$
\begin{aligned}
R(S) &= n^* \left(\lambda_A^* \Delta_A + \lambda_B^* E[\Delta_B]\right) \\
&+ \theta \lambda_B^* n^* \min\left\{ \frac{1}{\lambda_{AB}^* + \theta \lambda_{BB}^*}, \frac{1}{n^*(\lambda_A^* + \theta \lambda_B^*)} \right\} \left(\lambda_{AB}^* \Delta_A + \lambda_{BB}^* E[\Delta_B]\right) \\
&+ \lambda_A^* n^* \min\left\{ \frac{1}{\lambda_{AA}^* + \theta \lambda_{BA}^*}, \frac{1}{n^*(\lambda_A^* + \theta \lambda_B^*)} \right\} \left(\lambda_{AA}^* \Delta_A + \lambda_{BA}^* E[\Delta_B]\right), \qquad \text{(AC-5)}
\end{aligned}
$$

and

$$
\begin{aligned}
W(S) &= \frac{n^*}{r} \left[\lambda_A^* \frac{\Delta_A}{r} - C_A(\lambda_A^*) + \lambda_B^* \frac{E[\Delta_B]}{r} - C_B(\lambda_B^*) - rk \right] \\
&+ \frac{\lambda_A^* n^*}{r} \left[\lambda_{AA}^* \frac{\Delta_A}{r} - C_A(\lambda_{AA}^*) + \lambda_{BA}^* \frac{E[\Delta_B]}{r} - C_B(\lambda_{BA}^*) \right] \\
&\qquad \cdot \min\left\{ \frac{1}{\lambda_{AA}^* + \theta \lambda_{BA}^*}, \frac{1}{n^*(\lambda_A^* + \theta \lambda_B^*)} \right\} \\
&+ \frac{\theta \lambda_B^* n^*}{r} \left[\lambda_{AB}^* \frac{\Delta_A}{r} - C_A(\lambda_{AB}^*) + \lambda_{BB}^* \frac{E[\Delta_B]}{r} - C_B(\lambda_{BB}^*) \right] \\
&\qquad \cdot \min\left\{ \frac{1}{\lambda_{AB}^* + \theta \lambda_{BB}^*}, \frac{1}{n^*(\lambda_A^* + \theta \lambda_B^*)} \right\}. \qquad \text{(AC-6)}
\end{aligned}
$$

Example 2. As in Example 1, Δ_B follows the uniform distribution on $[0,1]$, $C_A(\lambda_A) = \frac{1}{2}\lambda_A^2$ and $C_B(\lambda_B) = 5\lambda_B^2$. Then, from (AC-3),

$$\begin{cases} rk = \frac{1}{2}\lambda_A^2 + 5\lambda_B^2, \\ [r + n(\lambda_A + \theta\lambda_B)]\,\lambda_A = \Delta_A - rk + \frac{1}{2}\lambda_{AA}^2 + 5\lambda_{BA}^2, \\ [r + n(\lambda_A + \theta\lambda_B)]\,10\lambda_B = \theta\pi_B - \theta rk + \frac{1}{2}\theta\lambda_{AB}^2 + 5\theta\lambda_{BB}^2, \\ [r + n(\lambda_A + \theta\lambda_B)]\,(\lambda_{AA} + \lambda_A) = 2\Delta_A - rk, \\ [r + n(\lambda_A + \theta\lambda_B)]\,(10\lambda_{BA} + \theta\lambda_A) = \theta\pi_B + \theta\Delta_A - \theta rk, \\ [r + n(\lambda_A + \theta\lambda_B)]\,(\theta\lambda_{AB} + 10\lambda_B) = \theta\Delta_A + \theta\pi_B - \theta rk, \\ [r + n(\lambda_A + \theta\lambda_B)]\,(10\lambda_{BB} + 10\lambda_B) = 2\theta\pi_B - \theta rk. \end{cases} \tag{AC-7}$$

where $\theta = 1 - S$ and $\pi_B = (1 + S)/2$.

Assume $r = 0.08$, $k = 1.2$, $\Delta_A = 0.3$, and let S vary. By numerical computation, we can plot λ_A, λ_B, λ_{AA}, λ_{BA}, λ_{AB}, λ_{BB} and n as functions of $S \in [0, 0.3)$ in Figure 2 which show shapes similar to those in Figure 1, respectively.

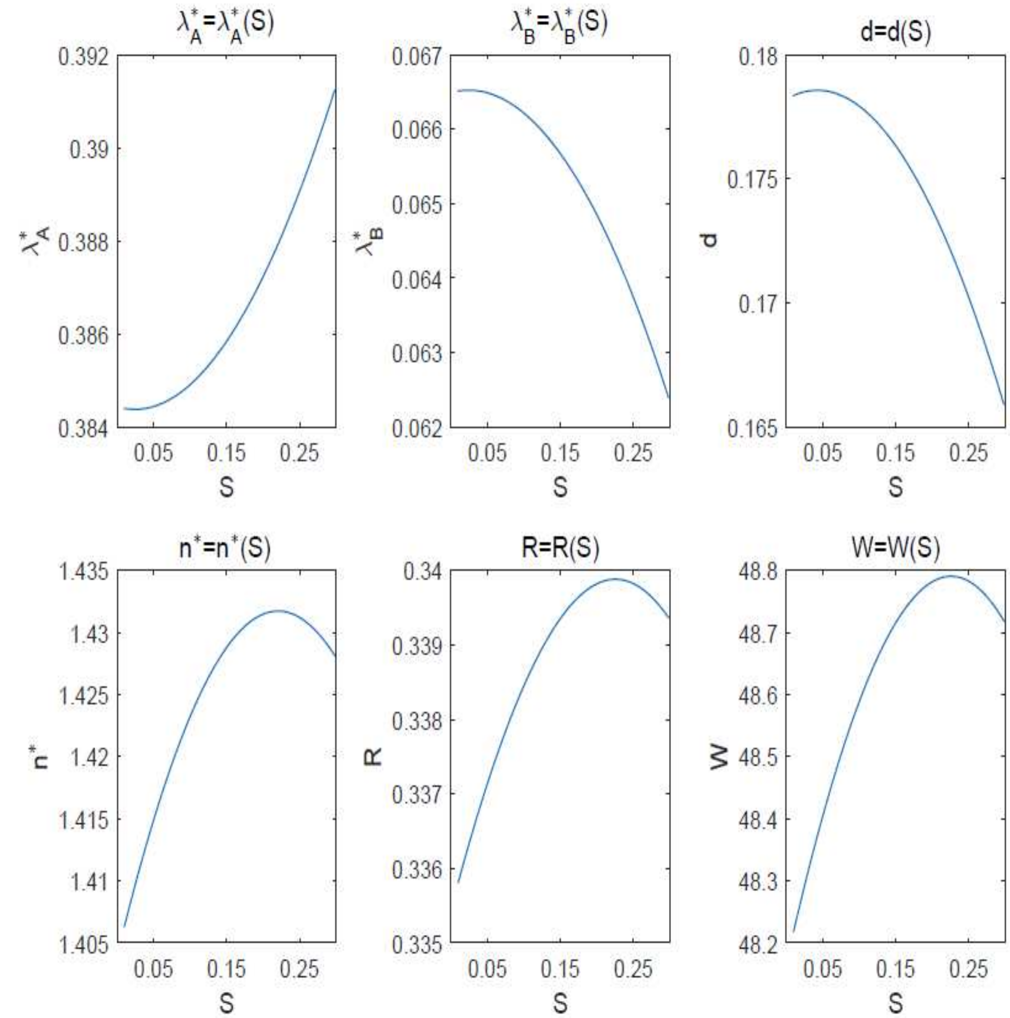

Figure 2: Example 2