Patent Protection for Compex Technologies



*Published Version:*
Vincenzo Denicolò (2022). Patent Protection for Compex Technologies. INTERNATIONAL JOURNAL OF INDUSTRIAL ORGANIZATION, 81, 1-20 [10.1016/j.ijindorg.2021.102811].

*Availability:*
This version is available at: https://hdl.handle.net/11585/894971 since: 2022-09-29

*Published:*
DOI: http://doi.org/10.1016/j.ijindorg.2021.102811

# Patent Protection for Complex Technologies*

Vincenzo Denicolò
University of Bologna and CEPR

Piercarlo Zanchettin
University of Nottingham

Decembre 13th, 2021

**Abstract**

We analyze patent protection with sequential and complementary innovation. We argue that in these cases the classic Nordhaus trade-off between innovation and static monopoly distortions is different from the case of isolated innovations. We parametrize the degree of innovation sequentiality and complementarity and show that the optimal level of patent protection increases with both. We also address the issue of the optimal division of profit among different innovators.

*Key words:* sequential innovation, complementarity, patent design, elasticity of the supply of inventions, division of profit

*JEL classification numbers:* O30, O40

*We thank three anonymous referees and the Associate Editor, Emilio Calvano, Giacomo Calzolari, Hugo Hopenhyan, Matthew Mitchell, Volker Nocke, Konrad Stahl and seminar audiencies at Helsinki, Lisbon, Mannheim, Tilburg and Chicago for useful comments and discussion. The usual disclaimer applies. E-mail addresses: vincenzo.denicolo@unibo.it (Vincenzo Denicolò) and piercarlo.zanchettin@nottingham.ac.uk (Piercarlo Zanchettin).

# 1 Introduction

In the modern economy, innovative products or processes often involve different patents, each protecting a separate piece of innovative knowledge. In some cases, basic discoveries open the way to subsequent improvements and applications, so innovation is *sequential*. In others, the distinct innovative components can be invented independently of each other, so innovation is *complementary*. In all of these cases, however, the aggregate value of the inventions is greater than the sum of their individual values. As a consequence, models of isolated, independent innovations are no longer applicable; technology is more "complex."

Many scholars have argued that when the technology becomes more complex,[1] the social costs of patent protection increase. One reason for this is that the fragmentation of intellectual property may create problems of coordination among the patent holders.[2] Furthermore, it has been argued that patents may impede the sharing of intermediate technological knowledge among innovative firms.[3] In the light of these issues, the conventional wisdom is that for complex technologies, patent protection should be weaker. Without questioning the relevance of these problems, this paper highlights a countervailing effect, which is of first-order magnitude but seems to have gone unnoticed so far. This effect implies that, all else equal, sequentiality and complementarity demand stronger and not weaker patent protection.

The new effect relates to the classic trade-off between innovation and monopoly deadweight losses. Since Nordhaus (1969), this trade-off has been at the centre of the analysis of optimal patent design for isolated innovations. The economic literature on sequential and complementary innovation, on the other hand, has mainly focused on other issues, such as the externalities across innovations and the division of the profit among the different innovators. Yet, sequentiality and complementarity may affect the Nordhaus trade-off and change its optimal resolution. That is precisely the

[1] An increase in complexity of patented innovations may be due not only to technological changes but also to institutional reforms. For example, the enaction of the Bayh-Dole Act in 1980, and the US Supreme Court decision in the *Diamond v. Chakrabarty* case in the same year, lengthened the chains of innovations protected by patents and thus increased complexity according to our usage of the word.

[2] Such coordination failures may generate transaction costs (Heller and Eisenberg, 1998; Galasso and Schankerman, 2015), pricing externalities (Lerner and Tirole, 2004), and greater scope for opportunistic behaviour (Farrell and Shapiro, 2008).

[3] In some models, this effect may be so strong that patent protection can actually stifle technological progress rather than spurring it (Bessen and Maskin, 2009; Fershtman and Markowich, 2010).

focus of the present paper.

The effect uncovered here can be intuitively explained as follows. With isolated innovations, the optimal level of patent protection is an increasing function of the elasticity of the supply of inventions. This elasticity is the percentage increase in in the probability of success $x$ associated with a one percent increase in R&D expenditure $X$, $\varepsilon = \frac{dx}{dX}\frac{X}{x}$. In simple models, the optimal level of protection is, in fact, directly proportional to the elasticity. Now consider the case of multiple, related innovations. For example, think of two innovations, 1 and 2, that are strictly complementary, meaning that the stand-alone value of each is nil but the aggregate value of both is positive. In this case, research is effectively successful only if both innovations are achieved, so the relevant probability of success (assuming statistical independence) is $x_1 \times x_2$. As a consequence, the relevant elasticity is now the *sum* of the individual elasticities, $\varepsilon_1 + \varepsilon_2$. This *compound-elasticity* effect is the ultimate reason why the level of patent protection should be higher than in the single innovation case. That is, in a nutshell, the message of this paper. The rest of the paper formalizes the above argument and extends it in various ways.

The compound-elasticity effect has been overlooked by the vast literature on sequential and complementary innovation because, as noted, that literature has almost invariably focused on issues other than the Nordhaus trade-off. In particular, the literature on sequential innovation has concentrated on the analysis of *forward* patent protection, i.e., the protection granted to the first inventor against the second one. This literature has either taken as given the level of *backward* protection, i.e., the protection against imitators, or conflated the two forms of protection in a single policy variable.[4] Similarly, the literature on complementary innovations has focused on the division of profit taking the overall level of protection as given.[5]

The only previous paper that distinctly addresses forward and backward protection is the classic article by Green and Scotchmer (1995) on sequential innovation. However, Green and Scotchmer drastically simplify the Nordhaus trade-off assuming that innovations can be achieved with probability one by sinking a fixed R&D investment. Since patents are distortionary, this implies that the optimal level of

[4] See, for instance, Scotchmer (1996), Matutes, Regibeau and Rockett (1996), O'Donoghue, Scotchmer and Thisse (1997), O'Donoghue (1998), Denicolò (2000), Denicolò and Zanchettin (2002), Hunt (2004), Hopenhayn, Llobet and Mitchell (2006), Mitchell and Zhang (2015), and Parra (2019).

[5] See, for instance, Shapiro (2007), Clark and Konrad (2008), Gilbert and Katz (2011), Schmidt (2014), Choi and Gerlach (2018).

backward protection is simply the one that allows innovators to just cover their R&D costs. With two-stage innovation, this principle applies separately to the two innovators. However, Green and Scotchmer further assume that the policymaker cannot control the division of profit finely. As a result, when both inventors' costs are covered, at least one of them will inevitably obtain a positive rent. This wasteful over-remuneration, which inevitably arises with sequential innovations, implies that backward protection should be stronger than with isolated innovations.

This paper arrives at the same conclusion as Green and Scotchmer but for different reasons. To highlight the differences, we abstract from the waste-of-profit effect assuming that the division of profit can be fine tuned. On the other hand, we posit a smooth relationship between R&D investment and innovation. This assumption is necessary for a proper analysis of the Nordhaus trade-off, as Green and Scotchmer (1995) recognize in their discussion of endogenous R&D investment (p. 31).

The rest of the paper proceeds as follows. In Section 2, we present and analyze our baseline model of sequential innovation. We parametrize the degree of sequentiality and provide conditions under which both backward and forward protection should be strengthened as the degree of sequentiality increases. In Section 3, we show that our conclusions are robust to several extensions of the baseline model. Section 4 presents similar results for the case of complementary innovations. Section 5 summarizes and offers some final remarks. Proofs are collected in the Appendix.

# 2 Sequential innovation

In this and the following section, we analyze the case of sequential innovation. We start from the case of two innovations, 1 and 2. (In Section 3, we extend the analysis to the case of an infinite sequence of innovations.) Innovations are sequential in that innovation 2 cannot be searched for unless innovation 1 has been achieved.

## 2.1 Baseline model

In our baseline model, we follow Green and Scotchmer (1995) in assuming that innovators are specialized. That is, certain firms invest only in innovation 1, and others only in innovation 2.[6] This assumption implies that intellectual property is inevitably fragmented. However, we assume that patent holders can perfectly coordinate their

[6] This assumption will be relaxed in Section 3.

behaviour. This allows us to abstract from effects that have already been highlighted by previous analyses, such as business stealing, transaction costs, Cournot complements etc.[7]

### 2.1.1 R&D technology

We posit an *innovation production function*

$$x_i = F_i(X_i) \tag{1}$$

that relates the probability $x_i$ that innovation $i$ is achieved to the aggregate R&D expenditure $X_i$ targeted to that innovation. We assume that the function $F_i(X_i)$ is smooth, increasing and concave. Concavity means that the returns to R&D are decreasing at the industry level, for instance because the production of innovative knowledge requires inputs that are in fixed supply, such as talent, or the set of good ideas at any given point in time. We also assume that $F_i(0) = 0$.

When there is competition in research, we further assume that returns to R&D are constant at the firm level (while being decreasing at the industry level).[8] This implies that each individual firm's probability of getting the patent, conditional on the innovation being achieved, is equal to its share in the aggregate R&D investment.

Under these assumptions, the elasticity of the innovation production function, $\varepsilon_i \equiv \frac{F_i'(X_i)X_i}{x_i}$, is positive and lower than one.[9] In what follows, we shall often refer to the iso-elastic specification $x_i = \gamma_i X_i^{\varepsilon_i}$, where $\varepsilon_i \in (0, 1)$ is constant and $\gamma_i > 0$ is a parameter. In general, however, the elasticity may be variable. For example, when the timing of innovation is stochastic and follows a Poisson process with arrival rate $\kappa_i X_i$, where $\kappa_i > 0$ is a parameter, the discounting-adjusted probability of success is $x_i = \frac{\kappa_i X_i}{\kappa_i X_i + r}$,[10] where $r$ is the interest rate. In this case, the elasticity is decreasing in

[7] These factors should be included in a more complete analysis. At this stage, however, abstracting from them allows us to better focus on the effect of interest.

[8] Accordingly to Griliches (1990), at the firm level,

> in the major range of the data [...] there is little evidence for diminishing returns, at least in terms of patents per R&D dollar. That is not surprising, after all. If there were such diminishing returns, firms could split themselves into divisions or separate enterprises and escape them. (p. 1167)

One reason why returns to R&D may be constant at the firm level and decreasing at the industry level is the possibility that two or more firms may independently achieve the same innovation. In this case, the patent will be granted to one firm only.

[9] A large empirical literature has tried to estimate this elasticity. Estimates vary considerably from study to study, however: see Cohen (2010) for an excellent survey.

[10] In the Poisson model, the innovation eventually arrives with probability one, and the variable $x_i$ becomes the discount factor corresponding to the time lag to innovation, as in Denicolò (2000).

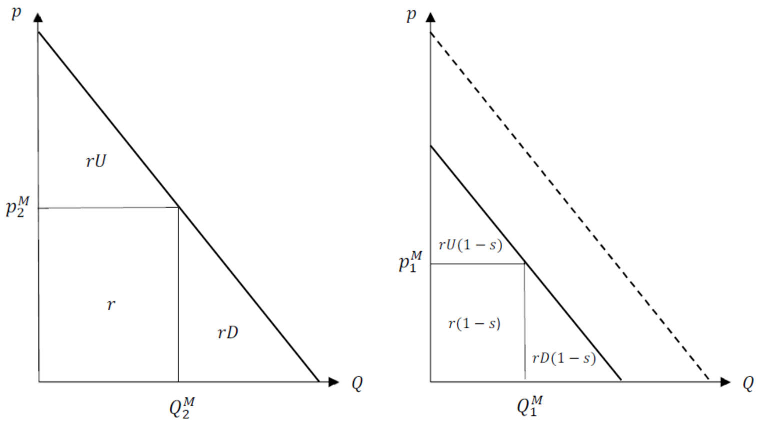


Figure 1: The components of the value of the innovations. The left-hand panel depicts the product market equilibrium when both innovations are available, the right-hand panel the equilibrium when only the first innovation has been achieved. The decreasing lines are the demand curves with one or both innovations in place. Marginal production costs are set to zero.

$X_i$. In the "buried treasure" model with no memory (Ross, 1983), on the contrary, the elasticity is increasing.[11]

### 2.1.2 Product market

We use a reduced-form model of the product market that is consistent with various different set-ups. For illustrative purposes, however, it may be useful to consider the example of Figure 1, where the first innovation creates a new product and the second innovation improves the product's quality, and hence shifts the demand curve upwards.

The left-hand panel represents the market equilibrium when both innovations have been achieved. With full patent protection, the price is set at the monopoly level $p_2^M$, as inventors coordinate their behavior at the pricing stage. This yields

[11] In this model, there are $N$ alternative R&D projects, only one of which is successful. Each project has a cost of $c$, so with no memory (i.e., assuming sampling with replacement) the probability of success is

$$x_i = 1 - \left(1 - \frac{1}{N}\right)^{\frac{X_i}{c}}.$$

It is then immediate to verify that the elasticity is increasing in $X_i$. Variants of the buried treasure model have been analyzed by Fershtman and Rubinstein (1997) and Chatterjee and Evans (2004).

a flow of monopoly profits that we normalize to $r$ (so that discounted monopoly profits are normalized to 1). Flow consumer surplus is denoted by $rU$ and the flow deadweight loss by $rD$. With no patent protection, on the other hand, price equals the unit cost, so profits vanish and flow consumer surplus becomes $r(1+U+D)$.

The right-hand panel shows the case where only the first innovation is available. Demand is lower, and so are all relevant payoffs. We denote by $s$ the share of the payoffs due to the second innovation, and by $(1-s)$ that due to the first. For simplicity, the parameters $D$ and $U$ are taken to be the same for both innovations.[12]

Generalizing this example, we assume that with no patent protection, the two innovations taken together raise social welfare by $1+U+D$, this benefit being reaped entirely by consumers. With full patent protection, in contrast, the social benefit is only $1+U$, as $D$ is lost due monopolistic distortions. Of this payoff, 1 is the innovators' profit and $U$ is the consumer surplus.[13] The first innovation accounts for a fraction $(1-s)$ of these payoffs.

Parameter $s$, which represents the share of total payoffs associated with the second innovation, will be our index of sequentiality. This index ranges between 0 and 1. When $s=0$, all of the value is attached to the first innovation and so we are back to the case of isolated inventions. When $s=1$, on the other hand, the first innovation has no direct value; it may be thought of as a pure research tool that is valuable only as it enables the search for the second innovation.[14]

### 2.1.3 Patent policy

In fact, the level of patent protection can take on intermediate values between complete protection or no protection at all. In our analysis, we shall treat the level of protection as a continuous policy variable and denote it by $\mu$. We normalize this variable in such a way that $\mu$ represents the inventors' total discounted profit if both innovations are achieved. Since the discounted profit with complete patent protection

[12]In the linear demand example, for instance, this property holds with $D=U=\frac{1}{2}$. However, the analysis would readily extend to the case where $D_1 \neq D_2$ and $U_1 \neq U_2$, as may happen with non-linear demand.

[13]The increase in consumer surplus when the patent is in force, $U$, is positive for drastic innovations, as in the example of Figure 1, but vanishes when innovations are non-drastic. (Innovations are non-drastic when competition from outsiders prevents innovators from charging the monopoly price. In this case, the innovators must engage in limit pricing and thus the quality-adjusted price falls only after the patent expires.) More generally, however, $U$ might include positive spillovers enjoyed by firms other than the innovators.

[14]Sometimes, $s$ is also referred to as the "option value" of the first innovation, as achieving the first innovation makes it possible to invest in the second one, which has an extra value of $s$.

is equal to 1, we have $\mu \in [0,1]$.[15]

There are various reasons why patent protection may be incomplete. Most obviously, the duration of the patent may be finite: for example, patent life is currently 20 years in most countries. Even before the patent expires, however, the breadth of protection may be limited. For example, the probability that the patent is granted and enforced, $\phi$, may be lower than 1. Denoting by $T$ the patent life, the expected discounted profit then is $\phi \int_0^T re^{-rt}dt = \phi(1-e^{-rT}) \equiv \mu$.[16]

Obviously, the level of patent protection affects also monopoly deadweight losses, which are $\phi \int_0^T rDe^{-rt}dt = \mu D$, and the surplus left to consumers, which is $(1-\mu)(1+D)+U$.[17] Assuming that the level of patent protection $\mu$ is the same for both innovations, all of these variables must be scaled down by the factor $(1-s)$ if only the first innovation is achieved.

#### 2.1.4 Sequentiality

With sequential innovation, the search for the second innovation can start only if the first one has been achieved. Accordingly, investment levels are chosen sequentially. In particular, investment in innovation 2 is made after uncertainty about innovation 1 is resolved, and thus it will be made only if innovation 1 has been successfully achieved. As a result, innovation 1 is achieved with probability $x_1$ and innovation 2 with probability $x_1x_2$.[18]

Besides determining the total profit obtained by the two innovators, patent policy

[15] This formalization implies that in the absence of patent protection, inventors make zero profits. The analysis can be extended to the case where secrecy is an alternative form of protection, as in Denicolò and Franzoni (2003). As these authors show, the possibility that the innovations may be kept secret does not change the analysis of optimal patent protection if secrecy entails the same social costs as patents. While monopoly distortions are likely similar, secrets may arguably stifle follow-up R&D in comparison with patents, thus creating an additional reason to strengthen patent protection on sequential innovations. This effect, however, is orthogonal to the compound-elasticity effect, and thus adding it to the analysis would not change our qualitative results.

[16] With other interpretations of patent breadth, however, it may be more difficult to conflate breadth and length in a single policy varable. In this case, there arises a non trivial problem of finding the optimal combination of patent length and breadth: see Gilbert and Shapiro (1990) for a classic analysis of this issue.

[17] Expected consumer surplus is

$$(1-\phi)\int_0^\infty r(1+U+D)e^{-rt}dt + \phi\left[\int_0^T rUe^{-rt}dt + \int_T^\infty r(1+U+D)e^{-rt}dt\right],$$

whence the expression in the text follows.

[18] While some time may pass from the arrival of the first innovation until that of the second, in what follows, for simplicity, we shall assume that such interim time interval is negligible. We shall therefore abstract from any payoff obtained in that interval.

determines also the division of the profit between them. We assume that innovator 1 gets the entire profit from the first innovation, i.e. $\mu(1-s)$, plus a fraction $\lambda$ of the profit from the second, $\mu s$. Innovator 2 gets the remaining share $(1-\lambda)$. The policy variable $\lambda \in [0,1]$ may be interpreted as the level of *forward* patent protection, i.e. the protection accorded to the first inventor against the second one.[19] In contrast, $\mu$ may be regarded as the level of *backward* protection, i.e. the protection against imitators.

### 2.1.5 R&D competition.

At this point, one can consider two variants of the model. In the first one, there is monopoly in the search for each innovation. That is, there are two firms, 1 and 2: firm 1 only can invest in innovation 1, and firm 2 in innovation 2. Firms are risk neutral and maximize their expected profits:

$$\pi_1 = x_1\mu(1-s) + x_1 x_2 \mu\lambda s - X_1 \tag{2}$$

and

$$\pi_2 = x_1\left[x_2\mu(1-\lambda)s - X_2\right]. \tag{3}$$

In the second variant, there is competition in R&D. For each innovation, a number of firms invest and race to innovate and obtain the patent. With constant returns to R&D at the firm level, the probability that each individual firm gets the patent, conditional on the innovation being achieved, is given by its share in the aggregate R&D investment in that innovation. Thus, the expected profits of two generic firms, investing respectively $Z_1$ in innovation 1 and $Z_2$ in innovation 2 are:

$$\pi_{z_1} = \frac{Z_1}{X_1}\left(x_1\mu(1-s) + x_1 x_2\mu\lambda s\right) - Z_1 \tag{2'}$$

and

$$\pi_{z_2} = x_1\left[\frac{Z_2}{X_2}x_2\mu(1-\lambda)s - Z_2\right]. \tag{3'}$$

In particular, we assume free entry in each R&D race. Under free entry, a zero-profit condition must then hold for each R&D race. Thus, the aggregate R&D investment levels are determined by the condition that $\pi_{z_1} = \pi_{z_2} = 0$.

[19] In practice, patent law affects the division of profit only indirectly, by determining what each patent holder is entitled to do unilaterally and hence the disagreement point in the bargaining process. However, taking $\lambda$ as a continuous variable simplifies the analysis and allows us to abstract from the difficulties of fine tuning the division of profit. As noted in the introduction, these difficulties are emphasized by Green and Scotchmer (1995).

Initially, we focus on the free-entry variant of the model. In Section 3, we shall show that similar results can be obtained when there is monopoly in research.

## 2.2 Equilibrium

With free-entry in R&D, the zero-profit conditions give:

$$x_1\mu(1-s) + x_1x_2\mu\lambda s = X_1 \tag{4}$$

and

$$x_2\mu(1-\lambda)s = X_2. \tag{5}$$

These conditions determine the equilibrium aggregate R&D investments, denoted by $X_i^*$,[20] and the corresponding probabilities of success, denoted by $x_i^*$. Notice that since firms invest in innovation 2 only if innovation 1 has been achieved, $X_2$ does not depend on $x_1$. The equilibrium is then found by proceeding backwards: condition (5) yields $X_2^*$ and hence $x_2^*$, which may then be plugged into (4) to obtain $X_1^*$. The existence of a positive solution can be guaranteed by the standard Inada conditions $\lim_{X_i \to 0} F_i'(X_i) = \infty$. Concavity of $F_i(X_i)$ guarantees uniqueness. To simplify the exposition, we assume interior solutions: $x_i^* < 1$.

This simple model delivers comparative statics results that are, for the most part, natural.

**Lemma 1** *Both $X_1^*$ and $X_2^*$ increase with the level of backward patent protection $\mu$. An increase in forward protection $\lambda$ decreases investment in the second innovation and, as long as $\lambda < 1 - \varepsilon_2$, increases investment in the first innovation.*

Intuitively, an increase in $\mu$ raises both innovators' profits and hence both R&D efforts.[21] As for $\lambda$, it is evident that stronger forward protection harms the second innovator. What is perhaps surprising is that it may stifle the first innovation, too. This happens when $\lambda > 1 - \varepsilon_2$, and the intuition is as follows. If the first inventor already gets a large share $\lambda$ of the "common" profit $\mu s$, a further increase in $\lambda$ may reduce the investment in the second innovation by so much that the first inventor is actually harmed. Clearly, this result by itself implies that it will never be optimal to

[20] As usual in models with constant returns at the firm level, the zero-profit conditions pin down the aggregate activity levels, but individual R&D investments are indeterminate.

[21] In addition to this direct effect, the impact of $\mu$ on $x_1^*$ involves also an indirect effect. That is, an increase in $\mu$ raises $x_2^*$ and hence the probability that the first innovator gets a share $\lambda$ of the second innovation's profits. Both effects are positive.

set $\lambda$ above $1-\varepsilon_2$, as in this region more forward protection would discourage both innovations.

### 2.3 Optimal backward protection

Having determined the model's equilibrium for any given $\mu$ and $\lambda$, we now turn to the optimal patent policy. We assume that the policymaker's objective is to maximize the expected discounted sum of consumer surplus and profit:

$$W = [1 + U + (1-\mu)D]\,[(1-s)x_1^* + s x_1^* x_2^*] - X_1^* - x_1^* X_2^*. \tag{6}$$

Under free entry in R&D, this is equivalent to maximizing consumer surplus only, as expected profits vanish. In this subsection, we take the level of forward protection $\lambda$ as given and focus on the optimal choice of backward protection $\mu$. In the next subsection, we shall optimize with respect to both policy variables, $\mu$ and $\lambda$.

**Proposition 1** *For any given level of forward protection $\lambda$, the optimal level of backward protection $\mu^*$ is higher when $s > 0$ than in the single innovation case ($s = 0$), provided that the elasticity $\varepsilon_1$ is non-increasing in $X_1$. When the elasticities $\varepsilon_1$ and $\varepsilon_2$ are constant, $\mu^*$ is a monotonically increasing function of $s$.*

In the proof of the proposition, it is shown that the optimal level of backward protection is implicitly given by the following condition:[22]

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \frac{\varepsilon_1}{1-\varepsilon_1} + S(1)\frac{\varepsilon_2}{1-\varepsilon_2} + S(\lambda)\frac{\varepsilon_1}{1-\varepsilon_1}\frac{\varepsilon_2}{1-\varepsilon_2}, \tag{7}$$

where

$$S(\lambda) \equiv \frac{\lambda s x_2^*}{(1-s)+\lambda s x_2^*}. \tag{8}$$

The variable $S(1)$ may be interpreted as an endogenous index of effective sequentiality. The index vanishes both when the second innovation has no value ($s = 0$), and when the second innovation is never achieved ($x_2^* = 0$). In both cases, we are effectively back to the single innovation framework. At the opposite extreme, the index equals one when the second innovation carries all of the value ($s = 1$). The variable $S(\lambda)$ is similar; it measures effective sequentiality from the viewpoint of the first innovator, which values the second innovation only as long as its share of profit $\lambda$ is positive.

[22] Expression (7) immediately implies that $\mu^* > 0$. This is intuitive, as social welfare vanishes when $\mu = 0$ (which in our simple model implies zero R&D investments). If condition (7) delivers a value of $\mu$ greater than 1, the optimal policy is to provide full patent protection ($\mu^* = 1$).

Since $S(\lambda)$ increases with $s$, the comparative statics results would follow immediately from (7) if the equilibrium R&D investments $x_1^*$ and $x_2^*$ were independent of $s$. However, they are not: as we increase $s$, $X_2$ increases and $X_1$ decreases. This complicates the analysis. The proof of the proposition shows that the sequentiality index $S(\lambda)$ is monotonic in $s$ even accounting for the endogeneity of the R&D investments. However, changes in $X_i$ may also affect the elasticities $\varepsilon_i$. This is why additional assumptions may be necessary to sign the comparative statics effects. In the first part of the proposition, for instance, the condition that $\varepsilon_1$ is non-increasing in $X_1$ guarantees that the effect of the fall in $X_1$ is non negative.[23]

The intuition behind Proposition 1 may not be immediate. To gain some insight, it may be useful to consider the marginal social costs and benefits of patent protection. To this end, assume for simplicity that $\lambda = 0$ and $U = 0$. Given the zero-profit conditions $X_1 = x_1\mu(1-s)$ and $X_2 = x_2\mu s$, social welfare reduces to:[24]

$$W = (1+D)(1-\mu)\left[(1-s)x_1^* + sx_1^*x_2^*\right]. \tag{9}$$

The first factor on the right-had side is independent of $\mu$, so the welfare effects of patent protection depend on the product of the last two factors. These two factors capture the two sides of the Nordhaus trade-off: $(1-\mu)$ represents the negative effect of stronger backward protection in terms of greater deadweight losses, whereas factor $[(1-s)x_1^* + sx_1^*x_2^*]$ captures the positive effect on the level of innovation. (Remember that both $x_1^*$ and $x_2^*$ increase with $\mu$.)

At the optimum, a small increase in $\mu$ must cause a percentage fall in the first factor equal to the percentage increase in the second. Now, a one percent increase in $\mu$ reduces $(1-\mu)$ by $\frac{\mu}{1-\mu}$ percent. This effect is independent of the degree of sequentiality $s$. The impact on the level of innovation $[(1-s)x_1^* + sx_1^*x_2^*]$ is more complex and can be calculated as follows. To begin with, note that as $\mu$ increases by 1%, if $x_1$ were constant $X_1$ would increase by 1% by the zero-profit condition.

[23] The condition is sufficient but not necessary. If $\varepsilon_1$ increases with $X_1$, the result is reversed only if this negative effect dominates the additional terms on the right-hand side of (7). While in general we cannot rule out this possibility, it should be noted that the extent to which the elasticity may be increasing is limited by the hypothesis of diminishing returns to R&D. Note also that while most of the empirical literature specifies the innovation production function as log-linear and hence implicitly assumes that the elasticity is constant, the few studies that allow for a variable elasticity find that the elasticity is decreasing (see e.g. Guo and Trivedi, 2002).

[24] Intuitively, expression (9) says that social welfare is equal to the expected social surplus obtained after the patent expires. Before the patent expires, $D$ is lost because of the monopoly distortions, and profits are "dissipated" in the patent races.

However, a 1% increase in $X_1$ raises $x_1$ by $\varepsilon_1$%, which by the zero-profit condition raises $X_1$ by a further $\varepsilon_1$%. This in turns raises $x_1$ by an extra $\varepsilon_1^2$%, and so on. Overall, $X_1$ increases by $1 + \varepsilon_1 + \varepsilon_1^2 + ... = \frac{1}{1-\varepsilon_1}$ percent, and thus $x_1^*$ increases by $\frac{\varepsilon_1}{1-\varepsilon_1}$ percent. By a similar argument, one sees that $x_2^*$ increases by $\frac{\varepsilon_2}{1-\varepsilon_2}$ percent. However, the product $x_1^* x_2^*$ increases by $\frac{\varepsilon_1}{1-\varepsilon_1} + \frac{\varepsilon_2}{1-\varepsilon_2}$ percent.[25] That is, the compound probability increases proportionally more than each individual probability. This is the *compound-elasticity* effect mentioned in the introduction.

To complete the argument, note that the percentage increase in $(1-s)x_1^* + sx_1^* x_2^*$ is a weighted average of $\frac{\varepsilon_1}{1-\varepsilon_1}$ and $\frac{\varepsilon_1}{1-\varepsilon_1} + \frac{\varepsilon_2}{1-\varepsilon_2}$, with the weight of the latter increasing with $s$. This means that as $s$ increases, the effect of a one-percent increase in $\mu$ on innovation becomes stronger. Since the negative effect on the deadweight loss is independent of $s$, and the two effects must exactly offset each other at the optimum, the optimal level of protection must increase when the degree of sequentiality $s$ increases.

Essentially, the above argument shows that the effectiveness of patent protection increases as the degree of sequentiality increases. However, other factors are also at work in the baseline model. First, the investment in the first innovation entails a positive externality on the second one. This *intertemporal externality* (to use the jargon of endogenous growth theory) is stronger, the greater is $s$. Second, increasing $s$ decreases the *ex ante* value of the innovations, $(1-s)x_1^* + sx_1^* x_2^*$. This is so because of sequentiality: for fixed levels of R&D investment, the second innovation occurs with a lower probability than the first one. This *devaluation* effect, too, is stronger, the greater is $s$. To isolate the compound-elasticity effect, in the next section we shall analyze variants of the baseline models where these additional effects are muted.

### 2.4 Optimal forward protection

Before doing that, however, consider the optimal choice of the level of forward protection $\lambda$. Since $\mu^*$ depends on $\lambda$ *via* the index of effective sequentiality $S(\lambda)$, the endogenization of $\lambda$ impacts also the optimal level of backward protection. Therefore, we also re-consider the comparative statics of $\mu^*$.

**Proposition 2** *When the elasticities $\varepsilon_1$ and $\varepsilon_2$ are constant, the optimal levels of forward protection $\lambda^*$ and backward protection $\mu^*$ are, respectively, weakly and strongly*

[25] When $\lambda > 0$, the increase is even bigger because the increase in $x_2$ raises the expected profit from innovation 1, and hence $x_1$. This additional effect is captured by the last term on the right-hand side of (7).

*increasing in the degree of sequentiality $s$ (both strictly increasing if $\lambda^*$ is positive).*

In the proof of the proposition, it is shown that the optimal level of forward patent protection is implicitly given by the following condition

$$\lambda^* = (1-\varepsilon_2)\,\frac{\frac{\varepsilon_1}{1-\varepsilon_1} - [1-S(1)]\,\frac{\varepsilon_2}{1-\varepsilon_2}}{\frac{\varepsilon_1}{1-\varepsilon_1} + S(1)\varepsilon_2}. \tag{10}$$

Expression (10) immediately implies that $\lambda^* < 1-\varepsilon_2$, confirming a conjecture made above. The level of forward protection is positive when

$$\frac{\varepsilon_1}{1-\varepsilon_1} > [1-S(1)]\,\frac{\varepsilon_2}{1-\varepsilon_2}. \tag{11}$$

When this inequality is reversed, we have a corner solution $\lambda^* = 0$. That is, the optimal level of forward protection vanishes if the elasticity of the supply of the first innovation is lower than that of the second one, and effective sequentiality $S(1)$ is low.

Besides re-affirming the results of Proposition 1 for the case where $\lambda$ is endogenous, Proposition 2 shows that $\mu$ and $\lambda$ are complementary policy tools: both should increase as the degree of sequentiality $s$ raises. We have already discussed the reasons why $\mu$ should increase with $s$. As for $\lambda$, the intuition is simple. The forward protection variable $\lambda$ allows to internalize, at least partially, the positive externality that the first innovator exerts on the second one. As $s$ increases, this intertemporal externality gets stronger, and thus we must raise $\lambda$ to internalize it more.[26]

# 3 Robustness

In this section, we analyze several extensions of the baseline model. To begin with, we consider alternative assumptions on who does the research. We then disentangle the effects at work in the baseline model, which have been informally discussed above, by developing models where some of the effects are muted. Finally, we consider the case of an infinite sequence of innovations.

[26] When the first innovation production function is less elastic than the second one ($\varepsilon_1 < \varepsilon_2$), we find a particularly interesting pattern. In this case, when $s$ is close to 0 inequality (11) is reversed and thus the optimal division of profit entails no forward protection (i.e., $\lambda^* = 0$). As the degree of sequentiality $s$ increases, it is initially optimal to increase only backward protection $\mu$. However, as $s$ increases further, the optimal level of forward protection $\lambda$ becomes positive. At the same time, $\mu$ should continue to increase. In other words, the first policy response to higher sequentiality should be to increase backward protection only. Forward protection should be used as an additional policy tool only for sufficiently high degrees of sequentiality.

## 3.1 Who does the research?

In the baseline model, we have assumed that there is free entry in each patent race. We now relax this assumption.

### 3.1.1 Specialized research monopolists

Consider the case of specialized research monopolists, where only one firm (firm 1) can invest in innovation 1, and another firm (firm 2) in innovation 2. The R&D investment levels are then determined by the first-order conditions:[27]

$$F_1'(X_1)\left[\mu(1-s)+x_2\mu\lambda s\right]=1 \tag{12}$$

and

$$F_2'(X_2)\mu(1-\lambda)s=1, \tag{13}$$

rather than by the zero-profit conditions. Everything else equal, the equilibrium levels of R&D investment $X_1^*$ and $X_2^*$ are lower: monopoly contracts R&D investment, just as it contracts output. However, with constant elasticities the comparative statics results are exactly the same as under free entry.[28]

Using consumer surplus as a welfare criterion,[29] the optimality conditions are also exactly the same as in the baseline model, i.e. (7) and (10). Therefore, Propositions 1 and 2 continue to hold.

However, the effective sequentiality index $S(\lambda)$ is now lower than in the baseline model, as equilibrium R&D investments are lower under monopoly than under free entry. As a result, the impact of an increase in $s$ on $\mu^*$ is less pronounced.

### 3.1.2 Repeated innovation

Another possible case is when only firm 1 can invest in innovation 1 but firm 1 may also compete with firm 2 in the search for innovation 2. This allows for the possibility of repeated innovation by the "technological leader."

As in the free-entry model, assume that the probability of success in the race for innovation 2 depends on the aggregate R&D expenditure, and that the probability that each firm $i=1,2$ gets the patent, conditional on the second innovation being

[27] Concavity of $F_i(X_i)$ ensures second order conditions.

[28] With variable elasticities, the terms $\frac{\varepsilon_i}{1-\varepsilon_i}$ are replaced by $\frac{\varepsilon_i}{1-\varepsilon_i(1-\eta_i)}$, where $\eta_i$ is the elasticity of $\varepsilon_i$ with respect to $X_i$. The sign of the comparative statics effects does not change.

[29] When profits are included in the social welfare function, formulas are more involved but the qualitative conclusions do not change.

discovered, is equal to its share in the total R&D investment in innovation 2, $\frac{X_{2i}}{X_2}$. The unconditional probability that firm $i$ wins the second patent will therefore be $x_{2i} = \frac{X_{2i}}{X_2} F_2(X_2)$.

Under these assumptions, it can be shown that both firms invest in the second innovation as long as $\lambda < 1 - \varepsilon_2$.[30] Using consumer surplus as the welfare criterion, the optimal level of backward protection is now given by:

$$\frac{\mu^*(1+D)}{[(1-\mu^*)(1+D)+U]} = \frac{\varepsilon_1}{1-\varepsilon_1} + S(1)\frac{\varepsilon_2}{1-\varepsilon_2} + \tilde{S}(\lambda)\frac{\varepsilon_1}{1-\varepsilon_1}\frac{\varepsilon_2}{1-\varepsilon_2}, \tag{14}$$

where

$$\tilde{S}(\lambda) = \frac{\left[\frac{1}{2}x_{21}^*(1-\varepsilon_2+\lambda)+\lambda x_{22}^*\right]s}{(1-s)+\left[\frac{1}{2}x_{21}^*(1-\varepsilon_2+\lambda)+\lambda x_{22}^*\right]s}. \tag{15}$$

A glance at the proof of Propositions 1 and 2 reveals that the only property of the sequentiality index used in the proofs is that it is increasing in $s$. The Appendix shows that $\tilde{S}(\lambda)$ does increase with $s$, so Propositions 1 and 2 continue to hold.

## 3.2 Disentangling the effects

In Section 2, we have argued that along with the compound-elasticity effect, the baseline model features also an intertemporal externality effect and a devaluation effect. In this subsection, we analyze variants of the baseline model where these additional effects are muted. This allows us to better identify the contribution of the compound-elasticity effect to making $\mu^*$ increase with $s$.

### 3.2.1 Internalizing the intertemporal externality

With sequential innovation, the first inventor exerts a positive dynamic externality on the second one. As long as $\lambda < 1$, this externality is not fully internalized and thus tends to create underinvestment in R&D – the more so, the greater the degree of sequentiality $s$. One may wonder that the impact of $s$ on the optimal level of backward protection depends on the need to correct for this externality.

To eliminate this possible effect, let us assume that all the research is done by one firm only, which can invest in both innovations. In this case, all externalities

[30] As in the free-entry model, setting $\lambda \geq 1 - \varepsilon_2$ is never optimal. If, nevertheless, inequality $\lambda \geq 1 - \varepsilon_2$ held, only the first inventor would invest in the race for the second innovation. The model would then become equivalent to the one considered in subsection 3.2.1 below. Note that in our model there is no Arrow effect, and thus as soon as $\lambda > 0$ firm 1 has a stronger incentive to invest in the second innovation than firm 2, as the latter must share its profit with the rival.

among innovators are fully internalized. Furthermore, the division of profit becomes irrelevant, so the only relevant policy tool is the level of backward protection $\mu$.

With monopoly in research, the innovator's expected profit is:

$$\pi = x_1\mu(1-s) - X_1 + x_1\left(x_2\mu s - X_2\right). \tag{16}$$

The first-order conditions for a maximum are:[31]

$$F_1'(X_1)\left[\mu(1-s) + x_2\mu s - X_2\right] = 1 \tag{17}$$

and

$$F_2'(X_2)\mu s = 1. \tag{18}$$

It is easy to confirm that the equilibrium R&D investments are still increasing in $\mu$.[32]

The optimal level of protection is now implicitly given by

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \frac{\varepsilon_1}{1-\varepsilon_1} + S(1)\frac{\varepsilon_2}{1-\varepsilon_2} + S(1-\varepsilon_2)\frac{\varepsilon_1}{1-\varepsilon_1}\frac{\varepsilon_2}{1-\varepsilon_2}. \tag{19}$$

The formula is the same as in the baseline model with $\lambda$ replaced by $1-\varepsilon_2$, so Proposition 1 still holds. The optimal level of backward protection, that is to say, increases with the degree of sequentiality $s$ even if the intertemporal externality effect is eliminated.

It may be interesting to analyze whether eliminating the intertemporal externality increases or decreases the magnitude of the impact of $s$ on $\mu^*$. With respect to the baseline model, there are two opposing effects. On the one hand, $\lambda$ is replaced by $1-\varepsilon_2$ in the third term on the right-hand side of (7). Since $\lambda < 1-\varepsilon_2$, and $S(\lambda)$ increases with $\lambda$ for given $x_2^*$, it follows that removing the intertemporal externality in itself magnifies the impact of $s$ on $\mu^*$.[33] On the other hand, however, $x_2^*$ is greater with free-entry in R&D, and the effective sequentiality index increases with $x_2^*$. In the

[31] As before, the second order conditions are easily verified.

[32] To be precise, the elasticities of the probabilities of success with respect to $\mu$ are

$$\frac{dx_2^*}{d\mu}\frac{\mu}{x_2^*} = \frac{\varepsilon_2}{1-\varepsilon_2},$$

just as in the baseline model case, and

$$\frac{dx_1^*}{d\mu}\frac{\mu}{x_1^*} = \frac{\varepsilon_1}{1-\varepsilon_1}\left[1 + S(1)(1-\varepsilon_2)\right].$$

[33] The intuitive reason for this is that when the intertemporal externality is fully internalized, a one percent increase in $\mu$ translates into a bigger percentage increase in the level of innovation. In other words, the key factor when the externality is internalized is not that the R&D investments increase but that they respond more strongly to an increase in the level of patent protection.

example pictured in Figure 2 below, this latter effect prevails, and thus the optimal level of backward protection is lower with a unique research monopolist than in the baseline model.[34]

### 3.2.2 Re-normalizing the value of innovations

In the baseline model, we have parametrized the degree of sequentiality $s$ in such a way that the *ex-post* value of the two innovations is constant as $s$ varies. However, the aggregate *ex-ante* value, which is proportional to $(1-s)x_1+sx_1x_2$, decreases with $s$, as the second innovation is achieved with a lower probability than the first one. Thus, when we increase $s$ the expected payoffs (both social and private) decrease. This may affect the way in which $\mu^*$ changes with $s$.

To eliminate this devaluation effect, we adjust the payoffs from the innovations by a factor $v$ that increases with $s$ in such a way that the *ex-ante* value

$$V = v\left[(1-s)x_1 + sx_1x_2\right] \tag{20}$$

stays constant. This requires that:[35]

$$\frac{dv}{ds} = \frac{(1-x_2)\,v}{(1-s)+sx_2} > 0. \tag{21}$$

In itself, the size of the innovations does not affect our optimality conditions (7) and (10). Thus, this re-normalization does not change Propositions 1 and 2. However, there is also an indirect effect: the effective sequentiality index $S(\lambda)$ increases with $x_2^*$, and $x_2^*$ increases with the size of innovations. As a result, the re-normalization raises the optimal level of backward protection $\mu^*$ for $s > 0$. In other words, the positive effect of $s$ on $\mu^*$ is amplified.

### 3.2.3 Quantifying the effects

Figure 2 illustrates the effects discussed so far and provides a quantitative assessment of their magnitude. The figure shows the optimal level of backward protection $\mu^*$ as a function of the degree of sequentiality $s$, assuming (when relevant) that forward protection is set to the optimal level $\lambda^*$. Optimal backward protection is lowest with

[34] With respect to the case of specialized research monopolists, however, $\mu^*$ is definitely higher with a single monopolist, because R&D investments are higher.

[35] Since the re-normalization depends on the equilibrium R&D investments, it is specific to each variant of the model.

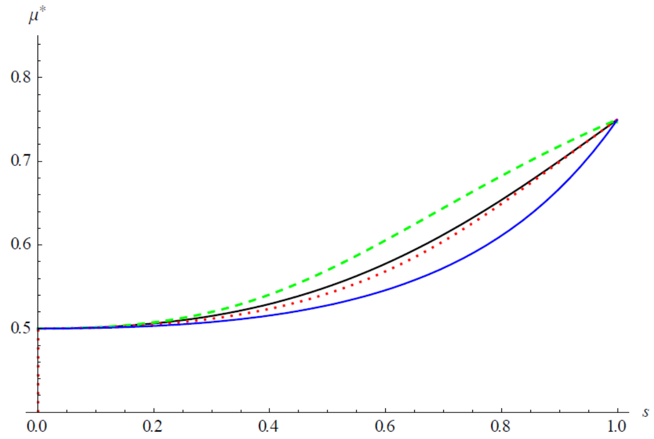


Figure 2: The optimal level of backward protection $\mu^*$ as a function of the degree of sequentiality $s$. Starting from below, the four curves represent, respectively, the case of specialized research monopolists (blue), the case of a single monopolist (red dotted), the baseline model (black), and the case where the value of the innovation is re-normalized (green dashed). The figure representes the case of iso-elastic production functions with $\varepsilon_1 = \varepsilon_2 = \frac{1}{2}$ and $\gamma_1 = \gamma_2 = 1$.

specialized research monopolists and highest under free entry, when the devaluation effect is eliminated. In all cases, however, $\mu^*$ increases with $s$.[36]

Figure 2 provides also an illustration of the magnitude of the effects uncovered in this paper. For the example considered, in the baseline model the optimal level of backward protection increases from $\mu^* = 0.529$ with $s = 0.4$ to $\mu^* = 0.578$ with $s = 0.6$. To get a sense of the magnitude of the change, with $r = 5\%$ the implied optimal patent length would increase from around 15 years to 17 years and 3 months. Comparing the extreme cases $s = 0$ and $s = 1$, the optimal patent length would double, from 14 to 28 years.

### 3.3 Infinite sequence of innovations

The baseline model considers a sequence of two innovations, as in the pioneering paper of Green and Scotchmer (1995). In this subsection, we show that our results extend also to models with an infinite sequence of innovations, as in O'Donoghue,

[36] The Mathematica file used to produce Figure 2 is available from the authors upon request.

Scotchmer and Thisse (1997), Hunt (2004).

For reasons of tractability, we assume stationarity, as nearly all models with an infinite sequence of innovations do. Time $t$ is continuous but is divided into periods of constant length $L$. The length of the period represents the time lag between successive innovations. In each period $\tau$, innovative firms make an aggregate investment $X_\tau$ and obtain innovation $\tau$ with probability $x_\tau = F(X_\tau)$. The innovation production function is assumed to be time invariant so as to guarantee stationarity. For the same reason, we assume that the profits generated by the innovations are stationary.

As in the two-stage model, sequentiality means not only that innovation $\tau + 1$ comes after innovation $\tau$ but also, and most importantly, that the former is enabled by the latter. Specifically, we assume that if innovation $\tau$ is achieved, then the innovation process can continue in period $\tau + 1$. If, on the other hand, innovation $\tau$ is not achieved in period $\tau$, then the innovative process stops forever. The analysis can be easily extended to the case where failure in period $t$ stops the innovative process only with a certain probability. However, as forcefully argued by Bessen and Maskin (2009), the assumption that this probability is positive is necessary for the model to be properly sequential.[37]

We now normalize to $r$ the perpetual flow of profit generated by each innovation under full patent protection. With patent protection set at level $\mu$, the flow of profit is then $r\mu$ and its present value, as of the beginning of period $\tau$, is $\mu$. In keeping with the baseline model, we assume that successive innovators do not compete with each other, so past innovators suffer no profit erosion as new innovations arrive.

In this framework, forward protection can be captured by assuming that innovator $\tau$ is entitled to a share $\lambda$ of the profits from innovation $\tau + 1$. Symmetrically, however, innovator $\tau$ must leave to innovator $\tau - 1$ a share $\lambda$ of the profits from innovation $\tau$. Because of a front-loading of profits effect (see Segal and Whinston, 2007), it is easy to see that the optimal level of forward protection is nil. In what follows, we shall therefore assume that $\lambda = 0$.

In this new framework, where all innovations are of the same size, we parametrize the degree of sequentiality by the interest rate $r$: the higher $r$, the lower the degree of sequentiality. This makes intuitive sense, as the discount rate determines the value of future innovations relative to the current one: the more heavily discounted the

[37] A model where the innovation process would continue even in case of failure would in fact be equivalent to one with isolated innovations.

future payoffs, the lower their value relative to the current payoffs.

#### 3.3.1 Equilibrium

With $\lambda = 0$, the expected profit from innovation $\tau$ is:

$$\pi_\tau = x_\tau \mu - X_\tau. \tag{22}$$

Assuming free entry, the equilibrium R&D investment in each period $\tau$, $x^*$, is then given by the zero-profit condition $\pi_\tau = 0$, or

$$F(X_\tau)\mu = X_\tau.^{38} \tag{23}$$

#### 3.3.2 Welfare

Denoting by $\delta \equiv e^{-rL}$ the period discount factor, and exploiting the zero-profit condition, social welfare can be recursively expressed as

$$W = x^* \left[(1-\mu)(1+D) + U\right] + x^* \delta W, \tag{24}$$

where the sequentiality assumption discussed above is captured by the fact that the continuation value $\delta W$ is obtained with probability $x^*$, i.e., only if the current inovation is achieved. Rearranging terms, the above expression can be rewritten as

$$W = \frac{x^*}{1-\delta x^*}\left[(1-\mu)(1+D) + U\right]. \tag{25}$$

Proceeding as in the two-period model, one finds that the optimal level of backward protection is now implicitly given by condition

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \frac{\varepsilon}{1-\varepsilon} + \bar{S}(r)\frac{\varepsilon}{1-\varepsilon} \tag{26}$$

where

$$\bar{S}(r) \equiv \frac{\delta x^*}{1-\delta x^*} > 0 \tag{27}$$

is the index of effective sequentiality for this model. The index is positive, implying that the optimal level of protection is higher than in the case of isolated innovations. Furthermore, the index $\bar{S}(r)$ is an increasing function of $\delta$, and hence a decreasing function of $r$. This implies that the level of protection increases with the degree of sequentiality, even in a model with an infinite sequence of innovations.

[38] From this, one can immediately derive the model's comparative statics:

$$\frac{dx^*}{d\mu}\frac{\mu}{x^*} = \frac{\varepsilon}{1-\varepsilon} > 0.$$

Similar results would hold under monopoly in research.

# 4 Complementary innovations

In this section, we consider the case where two complementary innovations can be achieved independently of each other. Accordingly, R&D investments are taken to be simultaneous rather then sequential.

## 4.1 Model assumptions

As in the sequential case, we normalize to 1 the aggregate profit from the two innovations under full patent protection, and we still denote by $U$ and $D$ the associated consumer surplus and deadweight losses. We now assume that a share $c$ of the payoffs is obtained only if both innovations are achieved. Of the remaining share $1-c$, a fraction $\beta_1$ is the stand-alone value of innovation 1 and the complementary fraction $\beta_2 = 1 - \beta_1$ that of innovation 2. The parameter $c$ is our measure of the degree of complementarity. When $c = 1$, innovations are strictly complementary; when instead $c = 0$, innovations are independent.

Unlike the sequential case, innovations can now be fully symmetric. This is so, in particular, if $\beta_1 = \beta_2 = \frac{1}{2}$ and $F_1(X_1) \equiv F_2(X_2)$.

Consistently with the case of sequential innovation, we assume that each innovator gets the full stand-alone profit, i.e., $\beta_1(1-c)\mu$ for innovator 1 and $\beta_2(1-c)\mu$ for innovator 2. Furthermore, innovator 1 gets a share $\lambda_1$ of the "common" profit $c\mu$, and innovator 2 the remaining share $\lambda_2 = 1 - \lambda_1$. The expected profits from each innovation therefore are:

$$\pi_i = \mu\beta_i(1-c)x_i + \lambda_i \mu c x_i x_j - X_i \tag{28}$$

With free entry in research, these expected profits must vanish.

## 4.2 Equilibrium and comparative statics

For any given patent policy $(\mu, \lambda_i)$, the zero-profit conditions $\pi_1 = \pi_2 = 0$ determine the level of R&D investments, $X_i^*$, and the equilibrium probabilities of success, $x_i^*$. The impact of patent policy on the equilibrium R&D investments can be found by implicit differentiation of the equilibrium conditions (28).

**Lemma 2** *Both $X_1^*$ and $X_2^*$ increase with the level of patent protection $\mu$ provided that*

$$\Omega \equiv (1-\varepsilon_i)(1-\varepsilon_j) - \varepsilon_i\varepsilon_j C_i\left(\frac{\lambda_j}{\beta_j}\right) C_j\left(\frac{\lambda_i}{\beta_i}\right) > 0, \tag{29}$$

*where*

$$C_i\left(\frac{\lambda_j}{\beta_j}\right) = \frac{\frac{\lambda_j}{\beta_j}cx_i^*}{(1-c) + \frac{\lambda_j}{\beta_j}cx_i^*}. \tag{30}$$

The index of effective complementarity $C_i\left(\frac{\lambda_j}{\beta_j}\right)$ is analogous to the indexes of effective sequentiality encountered in sections 2 and 3. It ranges from 0 when $c = 0$ to 1 when $c = 1$. However, differently from $c$, which is a purely exogenous measure of complementarity, the variables $C_i\left(\frac{\lambda_j}{\beta_j}\right)$ reflect also the equilibrium level of R&D investments and the division of profit among the innovators.

Condition $\Omega > 0$ may be viewed as a "stability" condition that ensures that firms do not get trapped in a zero-investment equilibrium in which investment in innovation $i$ vanishes for fear that innovation $j$ may not be achieved, and vice versa. The condition is always satisfied if the degree of complementarity is not too high.[39]

As for the division of profit $\lambda_i$, we have:

**Lemma 3** *As long as $\lambda_i$ is not too large, shifting profit from innovator $j$ to innovator $i$ increases $x_i^*$ and decreases $x_j^*$.*

When $\lambda_i$ is already very large (and thus $\lambda_j$ very low), however, a further increase in $\lambda_i$ may stifle both innovations. This effect is similar to the effect of increasing forward protection with sequential innovation and has the same intuitive explanation.

## 4.3 Optimal policy

Let us now turn to policy. Again, the policymaker's objective is to maximize the expected discounted sum of consumer surplus and profit. Under free entry, expected profits vanish and thus social welfare equals the expected consumer surplus from the innovations:

$$W = [(1-\mu)(1+D) + U]\left[\beta_1(1-c)x_1^* + \beta_2(1-c)x_2^* + cx_1^*x_2^*\right]. \tag{31}$$

### 4.3.1 Overall protection

We start from the optimal choice of patent protection $\mu$.

[39] When condition (29) fails, positive R&D investments may still be obtained if one innovation is targeted after the other in a pre-specified order, as in Biagi and Denicolò (2014). However, the model would then effectively become one of sequential innovations.

**Proposition 3** *For any given division of profit $\lambda_i$, the optimal level of patent protection $\mu^*$ is higher when innovation is strictly complementary ($c = 1$) than in the case of independent innovations ($c = 0$). If innovations are symmetric and the elasticities $\varepsilon_i$ are constant, $\mu^*$ is an increasing function of the degree of complementarity, $c$.*

In the proof of the proposition, it is shown that the optimal level of protection is implicitly given by the following condition:

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \sum_{\substack{i=1 \\ j\neq i}}^{2} \frac{C_i \left\{1-\varepsilon_j \left[1-C_j\left(\frac{\lambda_i}{\beta_i}\right)\right]\right\}}{(C_i+C_j-C_iC_j)\,\Omega}\varepsilon_i, \tag{32}$$

where $C_i = \frac{cx_i^*}{\beta_j(1-c)+cx_i^*} \equiv C_i\left(\frac{1}{\beta_j}\right)$.

As in the sequential case, the positive effect of $c$ on the optimal level of patent protection is due to the fact that complementarity raises the effective elasticity of supply of inventions. This can be seen most clearly in the symmetric case. Setting $\lambda_i = \frac{1}{2}$,[40] condition (32) reduces to:

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \frac{\varepsilon(1+C)}{1-\varepsilon(1+C)}, \tag{33}$$

where

$$C \equiv \frac{cx^*}{(1-c)+cx^*}. \tag{34}$$

This shows that the effective elasticity is now $\varepsilon(1+C)$ rather than $\varepsilon$ – yet another instance of the compound-elasticity effect. The underlying intuition is similar to that discussed above for the case of sequential innovation.

The complementarity indexes $C_i$ are increasing functions of $c$ and $x_i^*$. If the equilibrium R&D investments were independent of $c$, the comparative statics would then follow immediately. In fact, however, an increase in $c$ reduces both investments $X_i^*$, and hence both probabilities $x_i^*$. The reason for this is that when $c$ increases, a greater share of the value of the innovations is obtained with a lower probability. (This effect is similar to the devaluation effect that arises with sequential innovation.) The fall in $x_i^*$ reduces the indexes of effective complementarity, countering the direct positive effect of the increase in $c$. The proposition provides conditions under which the direct effect prevails.

[40] Setting $\lambda_i = \frac{1}{2}$ is indeed optimal in the symmetric case. This is obvious but can also be verified formally from condition (35) below.

#### 4.3.2 The division of profit

Finally, we turn to the issue of the division of profit. We have:

**Proposition 4** *If the elasticities $\varepsilon_i$ are constant, innovator $i$'s optimal share of the "common" profit $c\mu$ is an increasing function of $\varepsilon_i$ and a decreasing function of $\beta_i$.*

In the proof of the proposition, it is shown that the optimal division of profit is implicitly given by the following condition:[41]

$$\lambda_i^* = \frac{\varepsilon_i - \varepsilon_j + \varepsilon_j C_j}{\varepsilon_i C_i + \varepsilon_j C_j}. \tag{35}$$

Intuitively, $\lambda_i^*$ increases with $\varepsilon_i$ because a higher elasticity means that rewarding innovator $i$ becomes a more effective means for stimulating innovation. An increase in $\beta_i$, on the other hand, reduces the need to reward innovator $i$ with a share of the "common" profit because the stand-alone reward in itself is higher.

# 5 Conclusion

In this paper, we have analyzed the compound-elasticity effect that arises when innovation is sequential or complementary: in these cases, the effective elasticity of the supply of inventions is higher than for isolated innovations. A higher elasticity translates into a higher effectiveness of patent protection in stimulating R&D, and hence into a higher optimal level of protection. The compound-elasticity effect is robust: it operates in many different circumstances and does not depend on the specificities of the model used for the analysis.

In this paper, we have modelled patent policy in a highly stylized way. Accordingly, our analysis does not aim to make specific proposals for policy reforms. Its aim is, more modestly, to deliver a message of caution. The previous literature has highlighted various reasons why sequentiality and complementarity may raise the social costs of patent protection. This has created a conventional wisdom that patent protection should be weakened as the technology becomes more complex. This paper,

[41] It may be interesting to contrast (35) with the rule derived by Shapiro (2007), which is

$$\left(\frac{\lambda_i}{\lambda_j}\right)^2 = \frac{\frac{dx_i^*}{d\lambda_i}\frac{\lambda_i}{x_i^*}}{\frac{dx_j^*}{d\lambda_j}\frac{\lambda_j}{x_j^*}}.$$

Shapiro's rule is obtained by maximizing the aggregate expected profit rather than social welfare and is perhaps more opaque than (35) but has the same general flavour.

on the other hand, has shown that sequentiality and complementarity systematically raise the social benefits of patent protection. This countervailing effect is of first-order magnitude, suggesting more caution in drawing general conclusions about the impact of technological complexity on the optimal level of patent protection.

To derive more concrete implications for policy, the insights from our analysis must be incorporated into more highly structured models that account for specific institutional aspects of the patent system. This is an important task for future research. It would also be interesting to compare the size of the positive effect of greater technological complexity uncovered in this paper to the negative effects highlighted in the previous literature. This also would require a more comprehensive model and is left for future work.

## References


[1] Arrow, K. J. (1972). Economic welfare and the allocation of resources for invention, in *Readings in industrial economics* (pp. 219-236). Palgrave, London.

[2] Bessen, J. and E. Maskin (2009). Sequential innovations, patents, and imitation, RAND Journal of Economics, 40, 611-635.

[3] Biagi, A., and V. Denicolò (2014). Timing of discovery and the division of profit with complementary innovations, Journal of Economics & Management Strategy, 23, 89-102.

[4] Chatterjee, K., and R. Evans (2004). Rivals' search for buried treasure: competition and duplication in R&D. RAND Journal of Economics, 160-183.

[5] Choi, J. P. and H. Gerlach (2018). A model of patent trolls, International Economic Review, 59, 2065-2106.

[6] Clark, D. and K. Konrad (2008). Fragmented property rights and incentives for R&D, Management Science, 54, 969-981.

[7] Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance, in *Handbook of the Economics of Innovation*, Vol. 1, Amsterdam, North-Holland, 129-213.

[8] Denicolò, V. (2000). Two-stage patent races and patent policy, RAND Journal of Economics, 31, 488-501.

[9] Denicolò, V. (2007). Do patents overcompensate innovators?, Economic Policy, 22, 680-729.

[10] Denicolò, V., and L. A. Franzoni (2003). The contract theory of patents. International Review of Law and Economics, 23(4), 365-380.

[11] Denicolò, V. and P. Zanchettin (2002). How should forward patent protection be provided?, International Journal of Industrial Organization, 20, 801-827.

[12] Green, J. and S. Scotchmer (1995). On the division of profit with sequential innovation, RAND Journal of Economics, 26, 20-33.

[13] Farrell, J. and C. Shapiro (2008). How strong are weak patents?, American Economic Review, 98, 1347-1369.

[14] Fershtman, C. and A. Rubinstein (1997). A simple model of equilibrium in search procedures. Journal of Economic Theory, 72(2), 432-441.

[15] Fershtman, C. and S. Markovich (2010). Patents, imitation and licensing in an asymmetric dynamic R&D race, International Journal of Industrial Organization, 28, 113-126.

[16] Galasso, A. and M. Shankerman (2015). Patents and cumulative innovation: casual evidence from the courts, Quarterly Journal of Economics, 130, 317-369.

[17] Gilbert, R. and M. Katz (2011). Efficient division of profits from complementary innovations, International Journal of Industrial Organization, 29, 443-454.

[18] Gilbert, R. and C. Shapiro (1990). Optimal patent breadth and length, RAND Journal of Economics, 21, 106-112.

[19] Griliches, Z. (1990). Patent statistics as economic indicators: a survey, Journal of Economic Literature, 28, 1661–707.

[20] Guo, J.Q. and P.K. Trivedi (2002). Flexible parametric models for long-tailed patent count distributions, Oxford Bulletin of Economics and Statistics, 64, 63–82.

[21] Heller, M. and R. Eisenberg (1998). Can patents deter innovation? The anticommons in biomedical research, Science, 280, 698-701.

[22] Hopenhayn, H., Llobet, G. and M. Mitchell (2006). Rewarding sequential innovators: prizes, patents, and buyouts, Journal of Political Economy, 114, 1041-1068.

[23] Hunt, R. (2004). Patentability, industry structure, and innovation, Journal of Industrial Economics, 52, 401-425.

[24] Lerner, J. and J. Tirole (2004). Efficient patent pools, American Economic Review, 94, 691-711.

[25] Matutes, C., Regibeau, P., and K. Rockett (1996). Optimal patent design and the diffusion of innovations, RAND Journal of Economics, 27, 60-83.

[26] Mitchell, M. and Y. Zhang (2015). Shared patent rights and technological progress, International Economic Review, 56, 95-132.

[27] Nordhaus, W. (1969). *Invention, growth and welfare*, Cambridge, Mass., MIT Press.

[28] O'Donoghue, T. (1998). A patentability requirement for sequential innovation, RAND Journal of Economics, 29, 654-679.

[29] O'Donoghue, T., Scotchmer, S., and J. F. Thisse (1998). Patent breadth, patent life, and the pace of technological progress. Journal of Economics & Management Strategy, 7, 1-32.

[30] Parra, A. (2019). Sequential innovation, patent policy, and the dynamics of the replacement effect, RAND Journal of Economics, 50, 568-590.

[31] Ross, S. M. (1983). *Introduction to Stochastic Dynamic Programming*, Academic Press, New York.

[32] Scotchmer, S. (1996). Protecting early innovators: should second-generation products be patentable? RAND Journal of Economics, 27, 322-331.

[33] Schmidt, K. (2014), Complementary patents and market structure, Journal of Economics & Management Strategy, 23, 68-88.

[34] Segal, I., and M. Whinston,(2007). Antitrust in innovative industries, American Economic Review, 97, 1703-1730.

[35] Shapiro, C. (2007), Patent reform: Aligning reward and contribution, Innovation Policy and the Economy, 8, 111-156.

## Appendix

Proofs omitted in the text follow.

**Proof of Lemma 1.** By implicit differentiation of (5) w.r.t $\mu$ we obtain:

$$\frac{dx_2^*}{d\mu}\frac{\mu}{x_2^*} = \frac{\varepsilon_2}{1-\varepsilon_2} > 0. \tag{A1}$$

As for the impact of $\mu$ on $x_1^*$, differentiating (4) and (5) we have

$$\frac{dx_1^*}{d\mu}\frac{\mu}{x_1^*} = \frac{\varepsilon_1}{1-\varepsilon_1} + S(\lambda)\frac{\varepsilon_1}{1-\varepsilon_1}\frac{\varepsilon_2}{1-\varepsilon_2} > 0 \tag{A2}$$

where $S(\lambda)$ is given by expression (8) in the main text. The two terms on the right-had side correspond to the direct and indirect effect described in the main text, respectively.

Turning to the comparative statics with respect to $\lambda$, for the second innovation we have:

$$\frac{dx_2^*}{d\lambda}\frac{(1-\lambda)}{x_2^*} = -\frac{\varepsilon_2}{1-\varepsilon_2} < 0. \tag{A3}$$

As for the first innovation:

$$\frac{dx_1^*}{d\lambda}\frac{\lambda}{x_1^*} = S(\lambda)\frac{\varepsilon_1}{1-\varepsilon_1}\frac{1-\varepsilon_2-\lambda}{(1-\lambda)(1-\varepsilon_2)} \tag{A4}$$

This is positive if and only if $\lambda < 1-\varepsilon_2$. ■

**Proof of Proposition 1.** To begin with, we derive formula (7) in the main text. Using the zero-profit conditions (4) and (5), social welfare (6) reduces to:

$$W = [(1-\mu)(1+D)+U]\,[(1-s)x_1^* + sx_1^*x_2^*]. \tag{A5}$$

As noted, social welfare coincides with expected consumer surplus, as the profits from the innovations are entirely dissipated in the patent races.

The policymaker then maximizes (A5), keeping in mind that $x_1^*$ and $x_2^*$ are given by (4) and (5). To derive condition (7), we first rewrite (A5) as

$$W = V(\mu) \times \Xi(\mu,\lambda), \tag{A6}$$

where

$$V(\mu) \equiv [1+U+(1-\mu)D]$$

depends directly on $\mu$, and

$$\Xi(\mu, \lambda) \equiv [(1-s)x_1^* + sx_1^*x_2^*]$$

depends on $\mu$ and $\lambda$ only indirectly, through $x_1^*$ and $x_2^*$. Assuming an interior solution, the first order condition can then be written as:

$$-\frac{dV}{d\mu}\frac{\mu}{V} = \frac{d\Xi}{d\mu}\frac{\mu}{\Xi}.$$

To proceed, we calculate:

$$\frac{dV}{d\mu}\frac{\mu}{V} = -\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U}$$

and

$$\frac{d\Xi}{d\mu}\frac{\mu}{\Xi} = \frac{dx_1^*}{d\mu}\frac{\mu}{x_1^*} + S\frac{dx_2^*}{d\mu}\frac{\mu}{x_2^*}.$$

Using (A1) and (A2), condition (7), i.e.,

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \frac{\varepsilon_1}{1-\varepsilon_1} + S(1)\frac{\varepsilon_2}{1-\varepsilon_2} + S(\lambda)\frac{\varepsilon_1}{1-\varepsilon_1}\frac{\varepsilon_2}{1-\varepsilon_2}, \tag{7}$$

follows immediately.

The first part of the proposition follows immediately from (7). To see why, notice first of all that since the left-hand side of (7) is increasing in $\mu$, the solution $\mu^*$ is an increasing function of the right-hand side. Next, notice that when $s = 0$ both $S(1)$ and $S(\lambda)$ vanish, and thus (7) reduces to:

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \frac{\varepsilon_1}{1-\varepsilon_1}.$$

When $s$ is positive, on the other hand, two extra terms appear on the right-hand side of (7), and since they are both positive, they tend to increase $\mu^*$. On the other hand, however, when $s$ is positive $X_1$ is lower than when $s = 0$ (this follows from condition (4)). This is irrelevant if $\varepsilon_1$ is constant, but in general the fall in $X_1$ may affect the elasticity $\varepsilon_1$. The condition that $\varepsilon_1$ is non-increasing in $X_1$ guarantees that this additional effect is non negative.

To prove the second part of the proposition, since the left hand side of (7) is increasing in $\mu$, we just need to prove that the right-hand side is increasing in $s$. To this end, it suffices to prove that $S(\lambda)$ is increasing in $s$ for any $\lambda \in [0, 1]$. We calculate:

$$\frac{dS(\lambda)}{ds} = \frac{\lambda s(1-s)\frac{dx_2^*}{ds} + \lambda x_2^*}{[(1-s)+\lambda s x_2^*]^2}.$$

By implicit differentiation of (5) we get

$$\frac{dx_2^*}{ds}\frac{s}{x_2^*} = \frac{\varepsilon_2}{1-\varepsilon_2},$$

and plugging this into the preceding expression we finally have:

$$\frac{dS(\lambda)}{ds} = \lambda \frac{\frac{\varepsilon_2}{1-\varepsilon_2}(1-s)+1}{[(1-s)+\lambda s x_2^*]^2} x_2^* \geq 0$$

where the inequality is strict for $\lambda > 0$. ■

**Proof of Proposition 2.** To begin with, we derive formula (10) in the main text. Proceeding as in the proof of Proposition 1, notice that $\lambda$ enters (A6) only through the second term, $\Xi(\mu, \lambda)$. Assuming an interior solution, the first order condition can be written as:

$$\frac{d\Xi}{d\lambda}\frac{\lambda}{\Xi} = 0.$$

We then calculate:

$$\frac{d\Xi}{d\lambda}\frac{\lambda}{\Xi} = \frac{dx_1^*}{d\lambda}\frac{\lambda}{x_1^*} + S(1)\frac{dx_2^*}{d\lambda}\frac{1-\lambda}{x_2^*}\frac{\lambda}{1-\lambda}.$$

Using (A3) and (A4), condition (10) of the main text follows easily.

Next, we prove that when $\lambda^*$ is positive it is an increasing function of $s$. When $\lambda^* > 0$, condition (10) holds and by implicit differentiation we get

$$\frac{\partial \lambda^*}{\partial S(1)} = \frac{\varepsilon_2^2}{(1-\varepsilon_1)\left[\frac{\varepsilon_1}{1-\varepsilon_1} + S(1)\varepsilon_2\right]^2} > 0,$$

which means that $\lambda^*$ is an increasing function of the social sequentiality index $S(1)$. But we already know from the proof of Proposition 1 that $S(1)$ increases with the degree of sequentiality, $s$. This implies that when the solution is interior, the optimal degree of forward protection is a strictly increasing function of $s$. In turn, this implies that the solution $\lambda^*$ is, overall, weakly increasing in $s$.

Finally, we show that $\mu^*$ is increasing in $s$ when $\lambda$ is set at the optimal level $\lambda^*$. When inequality (12) is reversed, the optimal level of forward protection $\lambda^*$ is nil and hence does not depend on $s$. The result then follows directly from Proposition 1. Consider then the case in which inequality (12) holds and so we have an interior solution for $\lambda^*$. We have just shown that in this case $\lambda^*$ increases with $s$. To proceed, we must establish how changes in $\lambda$ affect $\mu^*$. We prove the following result:

*Result A.1. If*

$$\frac{\varepsilon_1}{1-\varepsilon_1} < (1-S(1))^2 \frac{\varepsilon_2}{1-\varepsilon_2} \tag{A7}$$

*then the optimal level of backward protection $\mu^*$ decreases with the level of forward protection $\lambda$. If instead inequality (A7) is reversed, then for any $0 < s < 1$ there exists a critical value of $\lambda$ , $\widehat{\lambda} \in (0,1)$, implicitly given by*

$$\widehat{\lambda} = 1 - \varepsilon_2 - \left[\frac{(1-s)+\lambda s x_2^*}{(1-s)+s x_2^*}\right]^2 \frac{\varepsilon_2}{\frac{\varepsilon_1}{1-\varepsilon_1}}, \tag{A8}$$

*such that the optimal level of backward protection $\mu^*$ increases with the level of forward protection $\lambda$ for $\lambda < \widehat{\lambda}$, while it decreases with $\lambda$ for $\lambda > \widehat{\lambda}$.*

*Proof of Result A.1.* As the left-hand side of (7) increases with $\mu$, $\frac{d\mu^*}{d\lambda}$ has the same sign as the derivative of the right-hand side of (7). That is:

$$\frac{d\mu^*}{d\lambda} \propto \frac{dS(\lambda)}{d\lambda}\frac{\varepsilon_1}{1-\varepsilon_1}\frac{\varepsilon_2}{1-\varepsilon_2} + \frac{dS}{d\lambda}\frac{\varepsilon_2}{1-\varepsilon_2},$$

where the symbol $\propto$ means "has the same sign as." Next, notice that

$$\frac{dS(\lambda)}{d\lambda} = \frac{s(1-s)}{[(1-s)+s\lambda x_2^*]^2}\left(x_2^* + \lambda\frac{dx_2^*}{d\lambda}\right); \quad \frac{dS(1)}{d\lambda} = \frac{s(1-s)}{[(1-s)+s\lambda x_2^*]^2}\frac{dx_2^*}{d\lambda}$$

Using the last two expressions and (A3), we then get:

$$\frac{d\mu^*}{d\lambda} \propto s(1-s)x_2^*\frac{\varepsilon_2}{1-\varepsilon_2}\frac{1}{[(1-s)+\lambda s x_2^*]^2}\times$$

$$\times\left\{\frac{\varepsilon_1}{1-\varepsilon_1}\left(1 - \frac{\lambda}{1-\lambda}\frac{\varepsilon_2}{1-\varepsilon_2}\right) - \frac{[(1-s)+\lambda s x_2^*]^2}{[(1-s)+s x_2^*]^2}\frac{1}{(1-\lambda)}\frac{\varepsilon_2}{1-\varepsilon_2}\right\},$$

which implies that $\frac{d\mu^*}{d\lambda}$ has the same sign as the term inside curly brackets. It is immediate to verify that this term is decreasing in $\lambda$ and vanishes when condition (A8) holds. Furthermore, at $\lambda = 0$ the derivative $\frac{d\mu^*}{d\lambda}$ has the same sign as

$$\frac{\varepsilon_1}{1-\varepsilon_1} - \frac{(1-s)^2}{[(1-s)+s x_2^*]^2}\frac{\varepsilon_2}{1-\varepsilon_2}$$

or, equivalently, as

$$\frac{\varepsilon_1}{1-\varepsilon_1} - [1 - S(1)]^2\frac{\varepsilon_2}{1-\varepsilon_2}.$$

Therefore, if inequality (A7) holds, then the derivative $\frac{d\mu^*}{d\lambda}$ is negative for any $\lambda \geq 0$. If instead inequality (A6) is reversed, then the derivative is positive for $\lambda < \widehat{\lambda}$ and negative for $\lambda > \widehat{\lambda}$. ■

Condition (12) is however reversed when inequality (A7) holds. Result A.1 then leaves us with $\lambda^* < \widehat{\lambda}$ as a sufficient condition for the proposition to be true. To show that the inequality $\lambda^* \leq \widehat{\lambda}$ indeed holds true, notice that condition (10) can equivalently be rewritten as

$$\left[\frac{(1-\lambda^*)}{\varepsilon_2} - 1\right] \frac{\varepsilon_1}{1-\varepsilon_1} = \frac{(1-s)+\lambda s x_2^*}{(1-s)+s x_2^*}$$

whereas condition (A8) may be rewritten as

$$\left[\frac{\left(1-\widehat{\lambda}\right)}{\varepsilon_2} - 1\right] \frac{\varepsilon_1}{1-\varepsilon_1} = \left[\frac{(1-s)+\lambda s x_2^*}{(1-s)+s x_2^*}\right]^2.$$

Since

$$\frac{(1-s)+\lambda^* s x_2^*}{(1-s)+s x_2^*} \leq 1,$$

it is clear that $(1-\lambda^*) \geq (1-\widehat{\lambda})$ and hence $\lambda^* \leq \widehat{\lambda}$. This completes the proof of the proposition. ■

**Omitted details for subsection 3.1.2.** Denoting by $x_{2j} \equiv \frac{X_{2j}}{X_2} x_2$ firm $j$'s probability of success in the second patent race, the expected profits of the two firms are:

$$\begin{aligned} \pi_1 &= x_1 \left[\mu(1-s) + x_{21}\mu s + x_{22}\mu\lambda s - X_{21}\right] - X_1 \\ \pi_2 &= x_1 \left[x_{22}\mu(1-\lambda)s - X_{22}\right]. \end{aligned}$$

Since R&D investments are chosen sequentially, to ensure subgame perfection we solve the model backwardly. In the second stage, the firms' best response functions are implicitly defined by the first order conditions:

$$\left[F_2'(X_2)\tfrac{X_{21}}{(X_{21}+X_{22})} + F_2(X_2)\tfrac{X_{22}}{(X_{21}+X_{22})^2}\right]\mu s + \left[F_2'(X_2)\tfrac{X_{22}}{(X_{21}+X_{22})} - F_2(X_2)\tfrac{X_{22}}{(X_{21}+X_{22})^2}\right]\lambda\mu s = 1$$

$$\left[F_2'(X_2)\tfrac{X_{22}}{(X_{21}+X_{22})} + F_2(X_2)\tfrac{X_{21}}{(X_{21}+X_{22})^2}\right]\mu s(1-\lambda) = 1.$$

The solution is independent of $x_1$ as firms invest in the second stage only if the first innovation has been achieved. Thus, the above conditions determine the R&D investments $X_{21}$, $X_{22}$ and $X_2$, and hence the associated probabilities $x_{21}$, $x_{22}$ and $x_2$. In the first stage, $X_1$ and $x_1$ will be determined by the first order condition:[42]

$$F_1'(X_1)\left[\mu(1-s) + \tfrac{X_{21}\mu s + X_{22}\lambda\mu s}{(X_{21}+X_{22})} - X_{21}\right] = 1.$$

[42]Notice that, whilst $x_{21} > 0$ for any admissible value of $\lambda$, $x_{22}^* > 0$ requires $\lambda < 1-\varepsilon_2$. Thus, when forward protecion becomes too strong, firm 2 stops investing in the second innovation and firm 1 enjoys complete monopoly in research.

With constant elasticities, i.e., $x_i = \gamma_i X_i^{\varepsilon_i}$, the model can be explicitly solved, yielding:[43]

$$\begin{aligned}
x_2^* &= \left[\frac{1}{2}\left(1-\lambda+\varepsilon_2\right)\mu s\right]^{\frac{\varepsilon_2}{1-\varepsilon_2}} \\
x_{22}^* &= \frac{1-\lambda-\varepsilon_2}{2\left(1-\lambda\right)\left(1-\varepsilon_2\right)}\left[\frac{1}{2}\left(1-\lambda+\varepsilon_2\right)\mu s\right]^{\frac{\varepsilon_2}{1-\varepsilon_2}} \qquad (36) \\
x_{21}^* &= \frac{(1-\lambda)-(1-2\lambda)\varepsilon_2}{2\left(1-\lambda\right)\left(1-\varepsilon_2\right)}\left[\frac{1}{2}\left(1-\lambda+\varepsilon_2\right)\mu s\right]^{\frac{\varepsilon_2}{1-\varepsilon_2}} \qquad (37) \\
x_1^* &= \varepsilon_1^{\frac{\varepsilon_1}{1-\varepsilon_1}}\left[\mu(1-s)+\frac{1}{2}x_{21}^*\left(1-\varepsilon_2+\lambda\right)\mu s+x_{22}^*\mu\lambda s\right]^{\frac{\varepsilon_1}{1-\varepsilon_1}}
\end{aligned}$$

The derivation of (15) follows the same steps as in the proof of Proposition 1. To show that $\tilde{S}(\lambda)$ increases with $s$, it suffices to prove that $x_{22}^*$ and $x_{21}^*$ are increasing in $s$. This is immediate from (A.8) and (A.9). ■

**Proof of Lemma 2.** By implicit differentiation of the system of the equilibrium conditions we have:

$$\frac{dx_i^*}{d\mu}\frac{\mu}{x_i^*} = \varepsilon_i \frac{1-\varepsilon_j\left[1-C_j\left(\frac{\lambda_i}{\beta_i}\right)\right]}{(1-\varepsilon_i)(1-\varepsilon_j)-\varepsilon_i\varepsilon_j C_i\left(\frac{\lambda_j}{\beta_j}\right)C_j\left(\frac{\lambda_i}{\beta_i}\right)} > 0. \blacksquare \qquad \text{(A10)}$$

**Proof of Lemma 3.** The result follows by implicit differentiation of the system of the equilibrium conditions, which yields:

$$\frac{dx_i^*}{d\lambda_i}\frac{\lambda_i}{x_i^*} = \frac{\left(1-\varepsilon_j\right)C_j\left(\frac{\lambda_i}{\beta_i}\right)-\frac{\lambda_i}{\lambda_j}C_i\left(\frac{\lambda_j}{\beta_j}\right)C_j\left(\frac{\lambda_i}{\beta_i}\right)\varepsilon_j}{\left(1-\varepsilon_i\right)\left(1-\varepsilon_j\right)-\varepsilon_i\varepsilon_j C_i\left(\frac{\lambda_j}{\beta_j}\right)C_j\left(\frac{\lambda_i}{\beta_i}\right)}\varepsilon_i. \qquad \text{(A10')}$$

The derivative is positive as long as

$$\lambda_i < \frac{\left(1-\varepsilon_j\right)\lambda_j}{C_i\left(\frac{\lambda_j}{\beta_j}\right)\varepsilon_j}. \blacksquare$$

**Proof of Proposition 3.** Proceeding as in the proof of Proposition 1, we first rewrite social welfare (31) as

$$W = V(\mu) \times \Xi(\mu, \lambda),$$

where

$$V(\mu) \equiv [1 + U + (1-\mu)D]$$

[43] To save on notation, we set $\gamma_i = 1$.

as before, but now

$$\Xi(\mu, \lambda) \equiv \beta_1(1-c)x_1^* + \beta_2(1-c)x_2^* + cx_1^*x_2^*.$$

Assuming an interior solution, the optimal level of backward protection is implicitly given by the condition

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \sum_{\substack{i=1 \\ j\neq i}}^{2} \frac{C_i}{C_i + C_j - C_iC_j} \frac{dx_i^*}{d\mu} \frac{\mu}{x_i^*}.$$

Plugging (A10) into this expression, one obtains condition (32) in the main text.

To prove the first part of the proposition, note that when $c = 0$ condition (32) reduces to

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \sum_{\substack{i=1 \\ j\neq i}}^{2} \frac{\beta_i x_i^*}{\beta_i x_i^* + \beta_j x_j^*} \frac{\varepsilon_i}{1-\varepsilon_i}. \tag{A11}$$

The right-hand side is a weighted average of the elasticity terms $\frac{\varepsilon_i}{1-\varepsilon_i}$ for the two innovations, with weights reflecting their respective stand-alone values.

When $c = 1$, on the other hand, (32) becomes

$$\frac{\mu^*(1+D)}{(1-\mu^*)(1+D)+U} = \sum_{\substack{i=1 \\ j\neq i}}^{2} \frac{\varepsilon_i}{1-\varepsilon_i-\varepsilon_j}. \tag{A12}$$

It is immediate to verify that the right-hand side of (A12) is greater than that of (A11). Since the left-hand side is increasing in $\mu$, it follows that $\mu^*$ is greater when $c = 1$.

Turning to the second part of the proposition, note that (33) implies that the sign of $\frac{d\mu^*}{dc}$ is equal to the sign of $\frac{dC}{dc}$. We calculate:

$$\frac{dC}{dc} = \frac{x^* + \frac{dx^*}{dc}c(1-c)}{[(1-c)+cx^*]^2}.$$

Thus, the sign of $\frac{dC}{dc}$ equals the sign of the numerator:

$$N(x^*(c), c) = x^*(c) + \frac{dx^*}{dc}c(1-c) \equiv N(c). \tag{A13}$$

The derivative $\frac{dx^*}{dc}$ can be found by implicitly differentiation of the equilirium conditions (28). This yields:

$$\frac{dx^*}{dc} = -\frac{\varepsilon}{1-\varepsilon(1+C)} \frac{x^*(1-x^*)}{(1-c)+cx^*} < 0. \tag{A14}$$

To show that $N(c)$ is positive, notice first that it is a continuous and differentiable funcion of $c$ on its domain $[0,1]$. Furthermore, $N(0) = x^*(0)$ and $N(1) = x^*(1)$, which are strictly positive values as condition (28) guarantees interior equilibria in R&D. Thus, for $N(c)$ to ever be negative there must exist at least two values of $c$, say $0 \le \underline{c} < \overline{c} \le 1$, such that $N(\underline{c}) = N(\overline{c}) = 0$, while $\frac{dN(\underline{c})}{dc} < 0$ and $\frac{dN(\overline{c})}{dc} > 0$. We prove below that this can never be true. Specifically, we show that at any $c$ such that $N(c) = 0$, it must be $\frac{dN(c)}{dc} < 0$.

To show this, we first differentiate $N(c)$ with respect to $c$:

$$\frac{dN(c)}{dc} = 2(1-c)\frac{dx^*}{dc} + c(1-c)\frac{d^2x^*}{dc^2}, \tag{A15}$$

and we calculate

$$\begin{aligned}\frac{d^2x^*}{dc^2} &= \frac{-\varepsilon\left\{\left[(1-c)(1-2x^*) - cx^{*2}\right]\frac{dx^*}{dc} + x^*(1-x^*)^2\right\}}{[1-\varepsilon(1+C)]\left[(1-c)+cx^*\right]^2} + \\ &\quad + \frac{\varepsilon^2 x^*(1-x^*)}{[1-\varepsilon(1+C)]^2\left[(1-c)+cx^*\right]}\frac{dC}{dc}. \end{aligned} \tag{A16}$$

We must evaluate the derivative $\frac{dN(c)}{dc}$ at $N(c) = 0$. First of all, from (A14) we obtain:

$$\left.\frac{dx^*}{dc}\right|_{N(c)=0} = -\frac{x^*}{c(1-c)}. \tag{A17}$$

Next, noting that

$$\left.\frac{\varepsilon}{1-\varepsilon(1+C)}\right|_{N(c)=0} = \frac{(1-c)+cx^*}{c(1-c)(1-x^*)}$$

and that $N(c) = 0$ implies $\frac{dC}{dc} = 0$, from (A17) we obtain:

$$\left.\frac{d^2x^*}{dc^2}\right|_{N(c)=0} = \frac{x^*\left[(1-c)(1-2x^*) - cx^{*2} - c(1-c)\left(1-x^*\right)^2\right]}{c^2(1-c)^2(1-x^*)\left[(1-c)+cx^*\right]} \tag{A18}$$

Finally, using (A17) and (A18), the derivative (A15) becomes:

$$\left.\frac{dN(c)}{dc}\right|_{N(c)=0} = -\frac{x^*\left[(1-c)+cx^*\right]}{c\left(1-c\right)\left(1-x^*\right)} < 0.$$

This implies that $N(c)$ can never be negative for $c \in [0,1]$, completing the proof of the proposition. ■

**Proof of Proposition 4.** The first-order condition for social welfare maximization is

$$\frac{d\Xi}{d\lambda_i}\frac{\lambda_i}{\Xi} = 0,$$

which may be rewritten as

$$\sum_{\substack{i=1 \\ j\neq i}}^{2} \frac{C_i}{C_i + C_j - C_i C_j} \frac{dx_i^*}{d\lambda_i} \frac{\lambda_i}{x_i^*} = 0,$$

or

$$C_1 \frac{dx_1^*}{d\lambda_1} \frac{\lambda_1}{x_1^*} = -C_2 \frac{dx_2^*}{d\lambda_2} \frac{\lambda_2}{x_2^*}.$$

Using (A10'), this condition becomes

$$\varepsilon_1(1-\varepsilon_2) - \lambda C_1 \varepsilon_1 = \varepsilon_2(1-\varepsilon_1) - (1-\lambda) C_2 \varepsilon_2.$$

From this one easily gets

$$\lambda_i^* = \frac{\varepsilon_i - \varepsilon_j + \varepsilon_j C_j}{\varepsilon_i C_i + \varepsilon_j C_j}. \tag{35}$$

Now, consider the impact on $\lambda_i^*$ of an increase in the elasticity $\varepsilon_i$. We have

$$\begin{aligned} \frac{d\lambda_i^*}{d\varepsilon_i} &= \frac{(\varepsilon_i C_i + \varepsilon_j C_j) - C_i (\varepsilon_i - \varepsilon_j + \varepsilon_j C_j)}{(\varepsilon_i C_i + \varepsilon_j C_j)^2} \\ &= \frac{\varepsilon_j C_j (1-C) + \varepsilon_j C_i}{(\varepsilon_i C_i + \varepsilon_j C_j)^2} > 0. \end{aligned}$$

Next, consider the impact on $\lambda_i^*$ of an increase in the stand alone value of innovation $i$, $\beta_i$. Note that $x_i^*$ increases with $\beta_i$, and that an increase in $\beta_i$ reduces $\beta_j$ on a one-to-one basis. As a result, an increase in $\beta_i$ raises $C_i$ and reduces $C_j$.

Differentiating (35), we have

$$\frac{d\lambda_i^*}{d\beta_i} = \frac{\varepsilon_j \frac{dC_j}{d\beta_i} (\varepsilon_i C_i + \varepsilon_j C_j) - (\varepsilon_i - \varepsilon_j + \varepsilon_j C_j) \left( \varepsilon_i \frac{dC_i}{d\beta_i} + \varepsilon_j \frac{dC_j}{d\beta_i} \right)}{(\varepsilon_i C_i + \varepsilon_j C_j)^2}.$$

Rearranging terms we get

$$\frac{d\lambda_i^*}{d\beta_i} = -\frac{\lambda_i^* \varepsilon_i \frac{dC_i}{d\beta_i} - \lambda_j^* \varepsilon_j \frac{dC_j}{d\beta_i}}{(\varepsilon_i C_i + \varepsilon_j C_j)} < 0,$$

where the inequality follows from the fact that $\frac{dC_i}{d\beta_i} > 0$ and $\frac{dC_j}{d\beta_i} < 0$. ■