Dynamics of patenting research tools and the innovations that use them

Bonwoo Koo and Brian D. Wright

International Food Policy Research Institute and

University of California, Berkeley

b.koo@cgiar.org

wright@are.berkeley.edu

Preliminary Document: Please Do Not Quote

September 18, 2002

Abstract

This study is a positive analysis of the implications of a uniform-life patent system in a two-stage model where one patented innovation, a "research tool," enables another. We assume an inelastically-supplied, efficient first innovation, and that licensing has no transaction cost. We show that (i) a longer patent life can encourage earlier subsequent innovation when the relative cost of the innovation lies in an intermediate range, (ii) the patent life limitation acts as a means of commitment against excessive exploitation by the holder of the initial innovation, facilitating earlier commercialization under license, (iii) the uniformity of patent parameters, a great administrative advantage of patents, is less problematic than generally realized, when follow-on innovation is competitive (iv) social welfare under monopoly can dominate competition for certain parameter ranges, and (v) an efficient *ex ante* licensing agreement can improve social welfare but reduce the royalty revenue to the first innovator.

I. Introduction

The fruits of innovation frequently materialize as end products of a sequence of prior innovations. The patenting of research tools, products and processes that have value only as inputs to further innovations, has increased over the past few decades, particularly in biotechnology and information technology. In biological science, for example, there are now many patents on research tools (e.g., gene sequences, cloning tools and methods, and means of genetic transformation) that play a critical role in follow-on innovations. Similarly, many patented software and hardware innovations are useful only in creating further programs or products. Under current patent law, the product of a research tool can be patentable even if the idea of using the tool for follow-on innovations is obvious. So, when a significant new research tool is patented and becomes available for licensing, there may be many potential users ready to compete to find and patent newly enabled follow-on innovations if the incentives are right.

How does the current patent system perform in this context? This paper addresses this question. Unlike the literature on optimal patent design, this study is a positive analysis of the implications of a uniform-life patent system in the simplest kind of dynamic sequence when one patented innovation enables another. We consider the particular case in which the initial innovation (the "research tool") is an input essential to a follow-on patented innovation that in turn lowers the cost of producing a consumer good.

We ignore renewal fees and licensing transaction costs, and abstract from choice of scope by assuming scope is infinite for the first patent with respect to follow-on enabled innovations.[1] We

[1] Thus we consider enabled innovations that infringe the patent on research tool. Consider, for example, a patent that reads on DNA sequences modified by a specific technique that enhances expression of trait such as insect-resistance. Assume such a sequence is incorporated in the genome of a new, patented transgenic cotton plant

also assume that the first innovation is exogenously made, and that consumer demand is such that patenting does not affect consumption. Thus we mainly focus on the effects of patent life on the incentive for follow-on innovation. Application of the enabling innovation is assumed to be costless absent any patent claims, and feasible on or after the date of the first patent which is assumed to be coincident with the announcement and availability of the innovation. The cost of research relative to the value of the second innovation, the duration of commitment of those resources, the nature of competition, and the period of overhang (time till expiration) of the previous patent determine the date of introduction of the infringing follow-on patented innovation.

The issue in this context is of incentives for sequential innovation is inherently dynamic, but in most models time plays a minor role (Gallini, 1992; Green and Scotchmer, 1995; O'Donoghue, 1998; Denicolo, 2000; Hopenhayn and Mitchell, 2001). The focus is instead on the size and division of profit between two predetermined innovators, induced by varying degrees of patent breadth or scope. These papers assume either that the follow-on innovation occurs immediately after the first innovation or completely replaces the first patent.[2] When, as is typical of such models, there is only one, predetermined, follow-on innovator, the single potential infringer chooses to initiate innovation immediately upon patenting the first innovation or when the first patent expires. However, when firms compete for the follow-on innovation, the dynamics are richer and reveal insufficiently recognized robustness of the current patent system to variation in the relevant economic parameters.

---

variety that produces cotton at lower cost than previously possible. The latter is a follow-on enabled innovation, infringing a prior enabling patent.

In the dynamic context, simulations have been used to study the effects of patent policy, because it is in general difficult to derive analytic results.[3] However, by addressing the limit of equilibria when the duration of commitment of research resources goes to zero, a polar opposite to the infinite-commitment assumption of Loury (1979), we have been able to derive analytical results that allow endogenous choice of time of entry and model competitive rent-dissipating behavior of follow-on innovators.

We show that under competition, if the cost of research is sufficiently low, an increase in patent life advances follow-on innovation as long as patent life is less than a critical level $T^* > 0$. The optimal patent life is finite in general, even though our assumptions on final demand (in contrast to Nordhaus 1969; Gilbert and Shapiro 1990; Klemperer 1990, among others) rule out static deadweight loss from the patent monopoly. A finite patent life can encourage the search for follow-on innovations that would be blocked under infinite patent life. Furthermore, endogenous competitive adjustment of the timing of follow-on investment can significantly reduce the sensitivity of social welfare to uniform patent life. Indeed in this context the patentability of the enabling technology improves the operation of the uniform-life patent system by increasing the robustness of its welfare benefits to variations in the economic profitability of competitive follow-on innovation.

By contrast, if only a single "monopolist," identified to the first innovator *ex post*, can make the second innovation, the uniform patent system is less robust to variations in the relative cost of follow-on innovation. However, monopoly avoids the problem of dissipation of at least part of the

---

[2] An exception is the interesting model of innovation and imitation by Gallini (1992) who considered the case where the possibility of imitation within the life of the patent affects the optimal patent life and the patenting decision by an innovator.

[3] For a numerical analysis of a similar model under more general cost conditions, see Koo (1998).

potential rents that arises when innovation is competitive, and so may yield a higher social welfare when the cost of innovation is low and follow-on innovation begins as soon as the first innovation is patented. But when the relative cost is higher so that follow-on innovation is delayed, it is made sooner under license when innovation is competitive, and involves no rent dissipation, given the delay.

Naturally, if an *ex ante* exclusive licensing agreement between the first innovator and a maximally efficient second innovator is reached at no transactions cost, it can yield the (weakly) highest social welfare since the second innovation is always made (for long enough patent life) immediately after the first innovation, and there is no rent dissipation. But the first innovator will not choose to designate a follow-on innovator as the exclusive candidate for *ex ante* Nash bargaining when *ex post* competition is an option, unless research costs for the second innovation are so high as to induce substantial delay.

Our results are obtained in a two-stage innovation model with features that make it amenable to analytical (as distinct from numerical) assessment introduced in section 2. In that section, we also discuss the *ex post* licensing equilibrium and efficient royalty rate for the use of a patent by a successful second innovator. Assuming a first innovation has already been made and that is fully revealed in the patent, we analyze in section 3 the incentive for the second innovation in a competitive innovation race, and derive the optimal patent life as a function of the relative cost of the second innovation. Section 4 considers the case where only a single “monopolist” can participate in the race for the second innovation, as assumed in several important models. The implications of an *ex ante* licensing agreement on the first innovator’s return and the monopolistic second innovators incentives and on social welfare are also considered in this section. Concluding remarks follow in section 5.

II. A two-stage innovation model

II.1. Assumptions

We adopt a highly stylized model with polar assumptions that allow us to focus analytically on the issues that are of interest here. The second innovation builds upon a patented first innovation, and is patentable. However, the scope of the patent is infinite, so the second patent always infringes the first patent. The first innovation is a research tool that does not impart any direct externality (such as scientific knowledge) other than that it enables the search for the second innovation. Thus, in contrast to many sequential innovation models, the innovations are pure complements in production of the consumer good.

To focus on the incentive for the second innovation, we assume (except where noted below) initially no rent dissipation in the first innovation; it occurs costlessly and exogenously at date zero, determined by the random and exogenous arrival of the idea to the first innovator. We assume that potential second innovators maximize expected profits, and (except where noted below) that a successful first innovator cannot compete for the second innovation.[4] The first innovator must patent her innovation to receive licensing revenue from the second innovator. A simple example is an independent researcher who receives a patent for a novel DNA sequence but does not have the capacity to conduct the specialized research necessary for further stochastic innovation that could produce a patented, lower-cost perfect therapeutic substitute for an existing drug.

The market for the search for the second innovation is initially assumed to be perfectly competitive, with free entry. We assume instantaneous free access for research purposes to the first innovation upon patenting and publication, at zero lag and zero marginal cost.[5] Following the tradition established by Loury (1979) and Lee and Wilde (1980), we assume each firm in the race chooses a Poisson hazard rate for a successful innovation, and the sum of these rates is $\lambda > 0$. The marginal cost of λ is assumed to be constant at c,[6] ruling out the firm-level scale economies or diseconomies in research that help determine market structure in models of the tradition of Loury and Lee and Wilde.

To facilitate the analysis, we also make a technical assumption that firms in the race commit some fraction of their current research flow costs for a period Δ in advance, where Δ may be arbitrarily small.[7] An alternative assumption with equivalent implications is that there is a time lag Δ between a successful innovation outcome and the arrival of this news to research administrators including competitors, resulting in rent dissipation during this lag. The specification of Loury can be interpreted as representing the special case in which Δ is infinity.

The first innovation does not directly produce any consumer good. It indirectly enables a flow value of v per period, accruing as a result of a reduction in the cost of a consumer good after achievement of the second innovation. To remove the static inefficiency of the patent monopoly that justifies limitation of patent life in Nordhaus (1969), we assume a perfectly inelastic demand

---

[4] If she could compete efficiently, she would keep the first innovation secret and patent the second innovation as a single innovative package, and the issues addressed here would not arise.

[5] The first innovator has every incentive to grant research licenses in this model.

[6] This assumption of constant marginal cost greatly simplifies the analysis, and also maximizes the rent dissipation associated with a patent race. See Brooks et al. (1999) for a discussion of the equivalence of the approach used here to a game-theoretic analysis. The models by Green and Scotchmer (1995), Chang (1995), and Denicolo (2000) implicitly assume a constant cost of innovation.

[7] Without loss of generality, we assume the fraction to be unity here.

curve for the consumer good, with zero income effect.[8] To focus on the dynamics of the innovation process itself, we also assume the level of intertemporal demand is static, in contrast to Barzel (1968) who compares the competitive and optimal timing of innovation when demand follows an upward trend, in an early example of a "real option" model.

There is, by assumption, a countable infinity of independent approaches, equally promising *ex ante*, of attempting to achieve the second innovation, and competitive innovators are able to avoid, almost surely, following an approach that has been pursued by another. The patent cannot be "invented around" by producing a substitute innovation. Thus, an innovation patented at time zero is fully protected till the end of the patent life, T. Together with the assumption of complementary innovations, this implies that delay of the second innovation reduces expected profit to be shared by the two innovators and the first innovator is worse off.[9]

In the environment with atomistic, unidentified, indistinguishable, and competitive potential licensees, the *ex ante* exclusive licensing agreement addressed in the two-firm world of Green and Scotchmer (1995) may not be feasible.[10] Among the many types of *ex post* licensing arrangements plausible for commercialization of the second innovation, are an up-front fee and a per-unit royalty (equivalent in our model because consumer demand is perfectly inelastic). For simplicity of exposition, we assume a running royalty, $\alpha \in [0, 1]$, the share of the gross revenue flow transferred from the second to the first innovator until the expiration of the latter's patent. Potential second innovators can adjust the timing of their entry into the innovation race in

---

[8] Every innovation is non-drastic in this specification and consumer surplus accrues only after the second patent expires.

[9] If successive innovations are marketable substitutes and patent scope is finite, the second innovator might invent around the first patent, rendering the first patent obsolete (Gallini, 1992; O'Donoghue, Scotchmer, and Thisse, 1998; Hopenhayn and Mitchell, 2001). In such models, delay of the second innovation increases the first innovator's expected profit by extending her stream of revenue from her own innovation.

response to the level of the royalty payment. We assume the transaction costs of licensing and royalty collection are zero. Firms are risk-neutral, and the discount rate is r.

II.2. Licensing equilibrium

Suppose that the second innovation is achieved and (immediately) patented before the first patent expires. The first patent holder has the right to block commercialization of the second innovation until her patent, received at date zero, expires at date T. The holder of second patent, awarded at date t and expiring at date $t + T$, can similarly block commercialization of his innovation. Thus, the sequential innovation process is vulnerable to the "double-holdup" problem (Merges and Nelson, 1990). However, if the rents to be gained are positive, they constitute an incentive to arrange a licensing agreement. The surplus to be divided is the flow gross value v from the date of issue of the second patent (assumed identical to the date of discovery) until T, the date of expiration of the first patent. Assuming Nash bargaining with equal bargaining power, the solution is a 50-50 share (i.e., $\alpha = .5$) of v during the $(T – t)$ years before the first patent expires.

When the cost of developing the second innovation is more than one half of the gross present value of revenue, the first innovator would prefer to make a credible offer of a royalty rate lower than .5 to render the second innovation more attractive to potential innovators. However, an *ex ante* announcement of a royalty rate below .5, proclaimed unilaterally by the first innovator, is not credible, due to the assumed absence of commitment capacity against *ex post* renegotiation once the R&D investment for the second innovation is sunk. The *ex post* royalty rate is thus $\alpha = .5$, independent of the cost of achieving the second innovation.

---

[10] Even if it were feasible, it might not be consistent with the interests of the first innovator, as we indicate below.

When patent life is infinite, competitive innovators who anticipate a 50% share of revenue will not pursue the second innovation if the cost is more than half the gross present value of the innovation. However, even if the royalty rate α stays at its equilibrium level of .5 and the cost is between 50% and 100% of the gross present value of the innovation, a second innovation with positive net social value can be induced, with a lag, by restriction of patent life to some finite value $T \in R^{+}$.[11] It might seem obvious that the lag is exactly T years, but under competition the expected lag is in general less than T years, as shown below.

III. Competitive second innovation

III.1. Timing of the second innovation

Assume the first patent with life T is received at date zero. If the second innovation is made and patented t periods later, the present value of the revenue from the innovation, net of royalty, evaluated at date t, is

$$V(t{:}T) = (1-\alpha)\int_{0}^{T-t} ve^{-r\tau}d\tau + \int_{T-t}^{T} ve^{-r\tau}d\tau. \quad (1)$$

During the overhang period till the first patent expires, the successful second innovator receives the flow revenue v, net of the royalty fee to the first patent holder (the first term on the right hand side (RHS) of equation (1)). Beyond the overhang, full revenue flow accrues to him until his own patent expires (the second term). As innovation is delayed, the overhang period from the first patent is shortened and the revenue increases. Potential competitors for the second innovation decide on the timing of their entry into the innovation race by comparing the revenue in equation (1) with the cost of innovation.

---

[11] Other credible alternative could be a compulsory royalty rate set by the government, a possibility under the patent laws of many countries.

Lemma 1: Timing of the second innovation under competition

Assume that at date $t \geq 0$, the second innovation has not yet occurred and its net present value $V(t, T)$ exceeds the cost c. In the limit, as the cost commitment period Δ approaches zero, the second innovation is, with probability one, made instantaneously at date t.

Proof. When identical, atomistic firms compete for an innovation at date t, the industry-wide expected profit, under the assumption of Δ-period commitment, is

$$\Pi(t) = \int_0^\infty \left[V(\tau, T) - c - c\lambda(\tau)(1 - e^{-r\Delta}) / r\right]\lambda(\tau) e^{-(r+\lambda(\tau))\tau} d\tau .$$

The first term of the integrand is the expected revenue evaluated at date τ, which is, for finite T, increasing as the timing of innovation τ is delayed. The second term is the expected research cost incurred till a successful innovation, and the third term is the extra pre-commitment cost incurred during an additional period Δ–representing, for example, the time delay of news of a successful innovation.

Since innovators have the option of setting innovative effort, the integrand is non-negative in equilibrium. The zero profit condition implied by free entry and constant marginal cost of search ensures that

$$\left[V(t, T) - c - c\lambda_c(t)(1 - e^{-r\Delta}) / r\right]\lambda_c(t) = 0$$

for all $t \geq 0$. Thus, the aggregate flow of innovative effort at t, if positive, is

$$\lambda_c(t) = \frac{r\left[V(t, T) - c\right]}{c\left(1 - e^{-r\Delta}\right)} .$$

For a positive rent gross of commitment cost ($V(t, T) - c > 0$), $\lim_{\Delta \to 0} \lambda_c(t) = \infty$. ♦

Lemma 1 implies that at $t = t^*$ where $t^* = \inf\{t \in R^+ : V(t, T) \geq c\}$, competitive firms rush into the race and the innovation is, in the limit as Δ approaches zero, made instantaneously, almost surely.[12] Thus, the model includes both rent dissipation and instantaneous innovation, and furnishes analytically tractable results.

Given a first innovation is patented (and revealed) at t = 0, the timing of the second innovation is easily calculated using the result of lemma 1. To simplify the exposition, we normalize the marginal and average cost of λ as a fraction of gross social value of the innovation if achieved instantaneously; i.e., $c \equiv kv/r$, where $k \in [0, 1]$ is the coefficient of relative cost. We implicitly assume the almost-sure limit as Δ goes to zero in statements about the timing of and return to innovation below. The profit to the second innovator from innovation at time $t \geq 0$ is

$$\pi(t,T,k) = \frac{v}{r}\left[(1-\alpha) + \alpha e^{-r(T-t)} - e^{-rT} - k\right]. \tag{2}$$

If the constant search cost is lower than the revenue when a firm makes an innovation immediately after the first patent ($c < V(0, T)$, or $k < (1-\alpha)(1-e^{-rT})$), then the second innovation follows immediately at date t = 0, dissipating the follow-on innovators' net expected rent. On the other hand, if the search cost exceeds the revenue after the expiration of the first patent ($k > 1 - e^{-rT}$), the second innovation will never be made. For an intermediate range of cost ($(1-\alpha)(1-e^{-rT}) \leq k \leq 1 - e^{-rT}$), the second innovation is made within the first patent life at time t^* ($\leq T$) when the expected rent to the second innovator equals zero. These results are summarized in the following proposition derived from equation (2) and the zero-profit condition for the second innovation:

Proposition 1: Timing of the second innovation

---

[12] This derivation of a free-entry research equilibrium with, in the limit, an infinite hazard rate is similar to the idea of an infinite exploration rate in a model of exploitation of exhaustible resource. Arrow and Chang (1982) claim without proof that the resource exploration rate can, in the limit, be infinite when the exploration cost is linear.

In the limit as the cost commitment period Δ goes to zero, under competition the second innovation occurs (almost surely) t^* *years after the first innovation is patented, where*

$$t^*(T) = \begin{cases} 0 & \text{if } 0 \le k < k_2 \equiv (1-\alpha)(1-e^{-rT}) \\ -\frac{1}{r}\log\left(\frac{\alpha e^{-rT}}{e^{-rT} + k - (1-\alpha)}\right) & \text{if } k_2 \le k \le k_3 \equiv (1-e^{-rT}) \\ \infty & \text{if } k_3 < k \le 1 \end{cases} \quad . \qquad (3)$$

Figure 1 shows the (almost sure) timing of the second innovation under different levels of search cost, when patent life is fixed at *T*. In a competitive innovation race, the timing (the solid line) is a non-decreasing function of the normalized search cost. Within the range of $k_2 < k < (1-e^{-rT})$, the second innovation may occur under competition with a lag, but within the life of the first patent. It is patented and exploited under license until the first patent expires, and then royalty-free for the remainder of the life of its own patent. Only low-cost second innovations ($k \le k_2$ under the assumption of competition) occur immediately after the first innovation is patented, as is often assumed in sequential innovation models (e.g., Green and Scotchmer, 1995).

When the second innovation occurs with a lag, its timing is sensitive to patent life, as summarized in the following comparative static result derived from equation (3).

Proposition 2: Patent life and timing of the second innovation

If the second innovation occurs with a lag, the lag is advanced or delayed by an increase in patent life T, depending on whether the cost k is greater or less than the second innovator's share (1 – α) *in the revenue flow.*

A longer patent life lengthens both the stream of revenue from the second innovation itself and the period of royalty payment to the first patent holder. Existing studies on sequential innovations often focus on the latter effect and argue that stronger patent protection of an initial innovation discourages follow-on innovators. However, when the cost of innovation is low, the former effect dominates the latter and the revenue to the second innovator increases with a longer patent life. Consumers benefit from the advance in the stream of consumer surplus that starts after the second patent expires. Proposition 2 indicates that stronger patent protection in the form of a longer patent life can encourage earlier follow-on innovation.

III.2. Social welfare and optimal patent life

Social welfare from the innovation sequence under a competitive race, $S_c(t^*)$, is almost surely

$$S_c(t^*) = \frac{\alpha v}{r}\left(e^{-rt^*} - e^{-rT}\right) + \frac{v}{r}e^{-r(t^*+T)}. \tag{4}$$

The first term on the RHS of equation (4) is the rent transferred via royalty to the first patent holder from the start of the second innovation t^* till the end of the first patent T. We assume that the first innovation is achieved by a single innovator by serendipity; none of this rent is dissipated by rent-seeking competition. The second component of social welfare is the consumer surplus accruing after the expiration of the second patent, at $t^* + T$. The rent to the successful second innovator, net of royalty fee to the first patent holder, is reduced to zero through free entry at t^*. The level of social welfare for different ranges of cost is summarized as:

$$S_c(T) = \begin{cases} \dfrac{v\left[\alpha + (1-\alpha)e^{-rT}\right]}{r} & \text{if } 0 \le k < k_2 \equiv (1-\alpha)(1-e^{-rT}) \\ \dfrac{\alpha v e^{-rT}(1-k)}{r\left[e^{-rT} + k - (1-\alpha)\right]} & \text{if } k_2 \le k \le k_3 \equiv (1-e^{-rT}) \\ 0 & \text{if } k_3 < k \le 1 \end{cases} . \tag{5}$$

The solid line in figure 2 relates social welfare under competition to the cost of search when patent life is finite at T and licensing is negotiated *ex post*.[13] Within the range of $0 \le k < k_2$, the second innovation occurs immediately (in the almost sure limit) at $t = 0$ and potential rent to the

---

[13] The parameters used in this and following figures are, unless otherwise noted, as follows: $v = 1$, $r = .1$, $T = 15$, $\alpha = .5$.

second innovator is completely dissipated, so social welfare is locally insensitive to the cost of the innovation. As the cost level increases above k_2, social welfare gradually decreases due to not only the higher current cost of innovation but also the delay in the second innovation, as shown in the solid line in figure 1. Social welfare drops to zero when the cost is too high to induce the second innovation: i.e., $k > (1 - e^{-rT})$.

If licensing is not feasible, due, for example, to high transaction costs or antitrust law, the gains from advance in the second innovation are greatly reduced, though not eliminated. In general, competition between firms ensures that the second innovation is patented within the first patent period at $0 \le t^* \le T$ (if $0 \le k \le 1 - e^{-rT}$) even though by assumption it cannot be commercialized before the first patent expires. The solid gray line in figure 1 shows that the second innovation is delayed for all range of costs if licensing is infeasible. This delay results in substantial reduction of social welfare, especially for a low cost of innovation, as shown in the solid gray line in figure 2. No rent accrues to the first innovator due to the absence of a license, and the consumer surplus flow after the expiration of the second patent constitutes the only contribution to social welfare.

It is widely recognized that a major benefit of licensing is that it motivates the first innovation via the prospect of royalty revenue (Green and Scotchmer, 1995). The above results show that, even when the first innovation is unresponsive to incentives, as assumed in our model, a licensing agreement can increase social welfare by advancing the second innovation and the date of onset of consumer surplus flow. This point is reinforced when we consider the case where the first innovation is not patentable. This case is equivalent to a single stage model where the first innovation has been provided by public sector. With no royalty payable to the previous innovator, all the rent net of innovation cost is dissipated through competition for the innovation race. By allowing patent for the first innovation, part of the dissipated rent is captured as royalty revenue and social welfare increases, assuming that part of rent is not dissipated in the search for the first innovation.

The relation between the timing of innovation and social welfare under competition is summarized in figure 3. Measuring flows in current values for expositional purposes, figure 3 illustrates the size of rent, cost of innovation, and consumer surplus under different timings of innovation. When the cost is low so that the second innovation is immediate at $t = 0$ (case a), half of the revenue is transferred to the first innovator and the remaining rent net of cost is dissipated through competition. The flow of consumer surplus starts after the second patent expires at T. If the cost is intermediate so that the second innovation is delayed till $t^* < T$ (case b), part of the rent is transferred to the first innovator only till her patent expires at T. The rest of rent matches the size of innovation cost through the adjustment of the timing of investment. If the cost is even higher so that the second innovation is made when the first patent expires (case c), all the rent covers the cost of innovation and consumer surplus which starts at $2T$ is the sole generator of social welfare.

Proposition 3: Optimal patent life under a competitive innovation race

When the innovation race is competitive, there exists a finite patent life $T^*(k)$ that maximizes social welfare by inducing entry either immediately or upon expiration of the first patent, where

$$T^*(k) = \begin{cases} -\frac{1}{r}\log\left(\frac{1-\alpha-k}{1-\alpha}\right) \ \ \textit{and} \ \ t^* = 0 & \textit{if} \ \ 0 \le k \le (1-\alpha) \\ -\frac{1}{r}\log(1-k) \quad \textit{and} \ \ t^* = T & \textit{if} \ \ (1-\alpha) \le k \le 1 \end{cases}.$$

These results are obtained from differentiation of equation (5):

$$\frac{dS_c}{dT} = \begin{cases} -v(1-\alpha)e^{-rT} & if \;\; 0 \le k < k_2 \equiv (1-\alpha)(1-e^{-rT}) \\ \dfrac{\alpha v e^{-rT}(1-k)[(1-\alpha)-k]}{[e^{-rT}+k-(1-\alpha)]^2} & if \;\; k_2 \le k \le k_3 \equiv (1-e^{-rT}) \\ 0 & if \;\; k_3 < k \le 1 \end{cases} .$$

For a low cost ($0 \le k < k_2$), the sign of dS_c/dT is negative, implying that a shorter patent life increases social welfare. Since the innovation is immediate at t = 0, the flow of consumer surplus begins earlier when T is reduced. As patent life decreases the upper bound k_2 falls, and the minimum patent life that induces an immediate second innovation is that which equalizes the upper bound with k: $T^* = -\log[(1-\alpha-k)/(1-\alpha)]/r$. When the cost is in an intermediate range ($k_2 \le k \le k_3$), the sign of dS_c/dT depends on the level of k. For $k_2 \le k \le (1-\alpha)$, social welfare increases with a longer patent life since the second innovation is advanced and so is consumer surplus. However, the lower bound k_2 also increases with patent life and the optimal patent life is that which equalizes the lower bound with k. The second innovation occurs immediately after the first patent within this range of cost. On the other hand, for $(1-\alpha) < k \le k_3$, the second innovation occurs between t = 0 and t = T, and social welfare decreases with patent life, as does the upper bound k_3. The optimal patent life is that which equalizes the upper bound with k, so that the second innovation occurs when the initial patent expires at $t^* = T^*(k)$, with no rent dissipation.

♦

Figure 4 shows that the optimal life correspondence admits a finite optimal patent life as a function of k. Under the optimally chosen patent life $T^*(k)$, the second innovation is immediate (at t = 0) for $k < (1 - \alpha)$ and delayed until $T^*(k)$ for $k > (1 - \alpha)$. When $k = (1 - \alpha)$, there are two optimal patent lives, one of which is infinite. The optimum patent life for $k = (1 - \alpha)$ is $T^*(k) = \inf\{T \in R^+: V(T, T) \ge kv/r\}$. If the cost of innovation strictly exceeds $(1 - \alpha)$ and patent life is infinite, the second innovation will never be achieved. However, under a finite patent life, any second innovation that offers a positive profit stream at that patent life, net of innovation cost, will be pursued by competitive innovators at or before the end of the first patent. Timing of the second innovation in this case is consistent with a common modeling assumption that follow-on innovation occurs after expiration of the first innovation at T. Society is better off with a finite-life patent since it can induce the second innovation that would be economically infeasible were patent life infinite.

The optimal patent life in proposition 3 is based on our assumption that the timing of the first innovation is exogenous and the first innovation is achieved efficiently. Assume alternatively that the exogenous idea for the first innovation is revealed to all competitive potential first innovators simultaneously, and the marginal research cost is constant, as for the second innovation in this model. Then the expected rent from royalties is dissipated and all social welfare comes from consumer surplus. The optimal patent life is the minimum necessary to cover the cost of the first innovation, in expectation.[14] Further, if we recognized static inefficiency from monopoly power, the optimal patent life would be shorter than the level in figure 4.

Though figure 4 shows that optimal patent life is highly sensitive to the cost in the neighborhood of $(1 - \alpha)$, the actual level of social welfare is not so sensitive to the choice of patent life. Figure 5 illustrates the level of social welfare as a function of patent life under different values of cost k, similar to the graph by Nordhaus (1969, figure 5.6, p. 83).[15] When k is slightly lower than $(1 - \alpha)$, the patent life that maximizes social welfare is relatively long at $T_2 = 39.13$ years (for $k = 0.49$). As patent life increases from T_0, the second innovation is gradually advanced (proposition 2) and this positive effect dominates the negative effect of the longer life of the

---

[14] Note that a longer patent life cannot increase the first innovator's royalty income in this case.

follow-on patent on the date of initiation of the post-patent consumer surplus flow. When patent life reaches T_2, the second innovation is immediate at $t = 0$ and social welfare is maximized. Any patent life above T_2 results in rent dissipation without affecting the timing of the innovation. At least half the possible social welfare from a sub-optimally chosen patent life is attained as long as the patent life is longer than a minimum threshold level of T_0 (= 6.93 years under the parameterization of the example).

On the other hand, if k is slightly higher than $(1 - \alpha)$, the welfare-maximizing patent life becomes very short, $T_1 = 7.14$ years (for $k = 0.51$). The second innovation is (almost surely) made at the expiration of the first patent. An increase in patent life above T_1involves a tradeoff: tt advances innovation and increases first innovator's royalty receipts, but delays the onset of consumer surplus flow. Thus social welfare is not highly sensitive to variations in patent life above the optimum. In general, the loss from choosing a life longer than optimal is much less important than if the possibility of licensing were ruled out, as shown in the solid gray line in figure 5.

IV. Monopolistic second innovation

IV.1. Timing of the second innovation

Consider an alternative model in which only a single firm (by assumption, not the first innovator) can pursue the second innovation, as could be true if it is the sole owner of a "crucial complement" to the first innovation. We first focus on the case of *ex post* agreement, and then consider the implications of *ex ante* agreement which imposes higher informational requirements. Without any threat of competition, the monopolistic second innovator can delay his investment beyond the break-even time, to maximize his profit.

---

[15] In figure 4, we parameterize the cost of innovation as $k = (1 - \alpha) \pm \varepsilon$, where $\alpha = .5$ and $\varepsilon = .01$.

Lemma 2: Speed of the second innovation under monopoly with *ex post* licensing at $t = T$.

In the limit as the commitment period Δ approaches 0, a second innovation that has not previously achieved occurs immediately at t = T, or never. There is no rent dissipation.

Proof. The monopolist's expected profit at date T with Δ-period commitment is

$$\begin{aligned}\pi_m(T) &= \int_0^\infty \left[V(T,T) - c - c\lambda(1-e^{-r\Delta})/r\right]\lambda e^{-(r+\lambda)\tau}\, d\tau \\ &= \frac{\lambda}{r+\lambda}\left[V(T,T) - c - c\lambda(1-e^{-r\Delta})/r\right]\end{aligned}.$$

Maximization of $\pi_m(T)$ yields the monopolist's optimum hazard rate:

$$\lambda_m(T) = r\left[\sqrt{\frac{(V(T,T)-c)}{c(1-e^{-r\Delta})} + 1} - 1\right].$$

As the commitment period Δ goes to zero, the search rate explodes: $\lim_{\Delta \to 0} \lambda_m(T) = \infty$. In the limit, success of the innovation is immediate. The total cost of search becomes $c\left[\int_0^\infty \lambda_m e^{-(r+\lambda_m)\tau} d\tau + \lim_{\Delta\to 0}(1-e^{-r\Delta})\lambda_m / r\right] = c$, and there is no rent dissipation. In the almost sure limit as $\Delta \to 0$ at $t = T$, the monopolistic second innovator's rent is $\pi_m(T) = e^{-rT}\left[(1-e^{-rT})v/r - c\right]$. We infer from this result that instantaneous monopolistic discovery at $t = 0$ yields, almost surely, $\pi_m(0) = (1-\alpha)(1-e^{-rT})v/r - c$. When $k = 1-\alpha-e^{-rT}$ so that $c = (1-\alpha-e^{-rT})v/r$, then $\pi_m(0) = \pi_m(T) = \alpha e^{-rT} v/r$. Since, at this value of k, $\pi_m(t)$ declines with t for $t < T$, an interior choice of monopolistic entry date is ruled out. ♦

The timing of the second innovation under monopoly is summarized as:

$$t^m = \begin{cases} 0 & \text{if } 0 \le k < k_1 \equiv 1-\alpha-e^{-rT} \\ T & \text{if } k_1 \le k \le k_3 \equiv (1-e^{-rT}) \\ \infty & \text{if } k_3 < k \le 1 \end{cases}.$$

Figure 1 shows that the second innovation is (weakly) delayed under monopoly (the dashed line) relative to the competition (the solid line), and the timing difference is greatest when the search cost is in an intermediate range ($k_1 < k < k_2$). Within this range, the second innovation is made immediately at $t = 0$ under competition because the expected rent is positive and firms rush into the innovation race. Under monopoly, however, the second innovation is delayed till the first patent expires because the marginal benefit from delay is larger than the marginal benefit from immediate investment. As the cost increases from k_2, the second innovation is delayed somewhat even under competition due to higher cost of innovation, and the innovation timing difference between competition and *ex post* monopoly, for a given T, gradually vanishes.

IV.2. Optimal patent life

To calculate social welfare under monopoly, the monopolist's rent must be added to the consumer surplus and royalty fee transferred to the first patent holder: $S_m(t^m) = \pi_m(t^m) + S_c(t^m)$. Social welfare for different cost levels is summarized as:

$$S_m(T) = \begin{cases} \dfrac{v(1-k)}{r} & \text{if } 0 \le k < k_1 \equiv 1-\alpha-e^{-rT} \\ \dfrac{v(1-k)e^{-rT}}{r} & \text{if } k_1 \le k \le k_3 \equiv (1-e^{-rT}) \\ 0 & \text{if } k_3 < k \le 1 \end{cases} . \quad (6)$$

The dashed line in figure 2 shows social welfare under monopolistic follow-on innovation and *ex post* licensing, for a given finite patent life. As the cost increases from zero, social welfare decreases due to reduced monopoly rent associated with higher search cost. Social welfare is higher than the level under competition in this range of cost due to the absence of rent dissipation under monopoly. When the cost reaches k_1, the timing of the second innovation is suddenly postponed from $t = 0$ to $t = T$ and social welfare drops due to the delay of monopoly profits and consumer surplus and the disappearance of the royalty fee transferred to the first patent holder. Competition dominates monopoly in this case since the innovation is earlier and so does the initiation of consumer surplus flow. Social welfare drops further to zero, when the search cost is too high to induce any second innovation ($k_3 < k \leq 1$).

If the patent life is infinite, social welfare under competition is given by rectangular shape which stays at the level under the horizontal line at $\alpha v/r$ as long as $k < (1 - \alpha)$ in figure 2. Social welfare under monopoly starts at v/r for $k = 0$ and decrease to $\alpha v/r$ for $k < (1 - \alpha)$ and drops to zero for higher cost. Monopoly dominates competition in this case, though no second innovation is made for high cost innovation with $k > (1 - \alpha)$. Limiting patent life induces the second innovation that would not be possible under infinite patent life, and favors competition over monopoly for a wide range of cost.

Proposition 4: Optimal patent life under monopoly

Under a monopolistic innovation market, there exists a finite patent life $T^m(k)$ that maximizes social welfare, where

$$T^m(k) = \begin{cases} -\frac{1}{r}\log(1-\alpha-k) & if \ \ 0 \leq k \leq (1-\alpha) \\ -\frac{1}{r}\log(1-k) & if \ \ (1-\alpha) \leq k \leq 1 \end{cases}.$$

Proof. Differentiation of equation (6) for $k_1 \leq k \leq k_3$ yields

$$\frac{dS_m}{dT} = -v(1-k)e^{-rT}.$$

Since the sign of dS_m/dT is always negative, the optimal patent life is derived from the upper bound $k_3 \equiv 1 - e^{-rT}$ and the second innovation is made at $t = T$. However, if k is less than $(1 - \alpha)$, the second innovation can be induced immediately at $t = 0$ by a sufficient increase in patent life. Thus, the patent life that maximizes social welfare is derived from the lower bound

$k_1 \equiv 1-\alpha-e^{-rT}$ for $k_1 \leq k \leq (1-\alpha)$ and from the upper bound k_3 for $(1-\alpha) < k \leq k_3$. When the cost is very low ($0 \leq k < k_1$), social welfare is indifferent to patent life, and the minimum patent life that induces an immediate second innovation is $T^m = -\log(1-\alpha-k)/r$. Similarly, for a very high cost ($k_3 < k < 1$), social welfare is independent of patent life and the lower bound k_3 is used to derive the minimum optimal patent life. ♦

The dashed line in figure 4 illustrates that the optimal patent life is in general longer under monopoly than under competition, when the search cost is low ($k < (1 - \alpha)$). Within this range of cost, the second innovation is immediate at $t = 0$ under both market structures if the patent life is optimally chosen. However, under monopoly, rent dissipation caused by a competitive innovation race can be prevented, and social welfare is maximized for any patent life above the dashed line of figure 4. When the cost is greater than $(1 - \alpha)$, the optimal patent life is the same under both market structures since the second innovation is made at $t = T$ (the expiration of the first patent). The first innovator gets nothing, and consumer surplus is the sole component of social welfare for both market structures.

IV.3. Implications of an *ex ante* licensing agreement

We now assume that an *ex ante* licensing agreement is feasible between the first patent holder and the second monopolistic innovator, before any investment for the second innovation is made. The possibility of an *ex ante* agreement affects the bargaining position of each innovator and the division of rents. Above we showed that when the cost is relatively high ($k_1 \leq k \leq k_3$) and only *ex post* agreement is possible, the second innovation is delayed till the first patent expires (dashed line in figure 1) and the first innovator cannot capture royalty rent. Within this range of cost, if all cost information is common knowledge, the first innovator will wish to induce the second innovation (delayed till T under *ex post* bargaining) within her patent period by agreeing *ex ante* to share the cost of the second innovation.

The second innovator's threat point is the expected present value at t of the profit from innovation made at T (> t). Under Nash bargaining, the first innovator sets the royalty rate to share equally the increase in expected profit from making the second innovation at t = 0, rather than at $t = T$. Thus, as is the case of Green and Scotchmer (1995), an *ex ante* agreement induces a sequence of innovations which were economically unfeasible under an *ex post* agreement, if the cost of the second innovation is relatively high.

However, for a low cost innovation ($0 \leq k < k_1$), under an *ex post* agreement, the second innovation is immediate at t = 0 and the first innovator receives $\alpha v(1-e^{-rT})/r$ as royalty revenue. If an *ex ante* agreement becomes feasible, the second innovator's threat point is the present value of the profits from delay of his investment and patent filing until expiration of the first patent, $\pi_m(T)$, and the first innovator is in a weaker bargaining position. The first innovator can be worse off with the opportunity for an *ex ante* agreement due to the creation of the option to delay by the second innovator, protected by *ex ante* exclusivity. The first innovator has no incentive to identify the second innovator with an *ex ante* agreement, if the cost of the second innovation is expected to be low.[16]

Society is always better off, of course, with *ex ante* negotiation, as shown by the dotted line in figure 2. The second innovation is immediately made after the first patent for all range of cost less than $(1-e^{-rT})$, and its social welfare dominates all other cases. For a low cost innovation,

[16] If the cost of the innovation is very low compared to the value, the second innovator may be worse off by delaying innovation till the first patent expires. Then the second innovation will be made regardless of the timing of licensing agreement.

rent dissipation associated with a competitive innovation race is prevented without delaying the second innovation. For a high search cost, an *ex ante* monopolistic agreement retains the advantage of inducing the second innovation at $t = 0$ rather than later. But if only *ex post* agreement is possible, competition dominates monopoly in follow-on research, for cost above k_1.

V. Conclusion

Our analysis indicates that the operation of the patent system in a world of sequential innovation has several positive features quite distinct from those emphasized in standard static models. We assumed a highly stylized two-stage model with an inelastic first innovation (the research tool) and constant marginal cost of innovation, but we allowed follow-in innovations using the tool to occur within the previous patent period, and for licensing agreement for commercialization, features often neglected in other models. By introducing the concept of an arbitrarily brief commitment period Δ for R&D investment, which facilitates analysis of rent-seeking dynamics, we show that a longer patent life does not hinder subsequent innovation when the relative cost of the innovation lies in an intermediate range. If the gross profit margin is modest, the limit on patent life can induce subsequent innovation that would be precluded under infinite patent life. The patent life limitation acts as a means of (partial) commitment against excessive exploitation by the holder of the innovation essential to follow-on innovation, facilitating earlier commercialization under license. Patent life optimally adjusted for profitability of the innovation induces either immediate patenting of the follow-on innovation, or patenting at the expiration of the term of the research-tool patent.

The analysis also reveals that, under competition, the uniformity of patent parameters, a great administrative advantage of patents, is less problematic than generally realized. The welfare gains from patenting are fairly robust to differences in profitability. The patent system allows competitive follow-on firms to adjust the timing of their investment to mitigate the distortion caused by a patent. As long as the patent life is set long enough to induce follow-on innovations, the cost of a sub-optimal patent life consists of dissipation of the rents net of royalties for use of the research tool, and of some delay of the surplus flow for intermediate cost innovations. Only very low-profitability innovations will be delayed for a term as long as the life of the patent.

Social welfare is less robust to cost differences under monopoly. Under *ex post* licensing the high-cost second innovation comes later, and social welfare is lower, than the level under competition. However, if the second innovation is cheap, social welfare under monopoly can dominate competition because competitive rent dissipation can be completely, rather than only partially, avoided.

Achieving an efficient *ex ante* licensing agreement, if feasible, may improve social welfare for a relatively high-cost case by advancing the second innovation. However, the possibility of an *ex ante* agreement may increase the bargaining power of the second innovator. Indeed, the first innovator can lose by identifying the follow-on innovator *ex ante* if the innovation is expected to be highly profitable and *ex post* competition is an option.

The assumptions we have made here are extreme in order to facilitate the analysis and exposition. By assuming away the consumer surplus during the patent protection (as under a 'drastic' innovation), we might have excessively favored the monopolistic environments. Had we assumed a positive stream of consumer surplus during the patent protection, rent dissipation from competition might not be so important as the timing of innovation itself, and competition is more likely to be preferred to monopoly. Similarly, if we assumed firms are different in their efficiency of innovation, a competitive market should be superior in identifying the most efficient innovator. We also did not explicitly consider the incentive and timing of the first innovation and we ruled out rent dissipation in the first innovation. The results presented here on timing are identical if there is rent dissipation in the first innovation race, but the optimal patent life tends to be shorter. Moreover, we exaggerated the benefits of licensing by ignoring positive transaction cost associated with licensing, which can preclude innovation altogether in severe enough cases (Heller and Eisenberg 1998).

Based on numerical results not reported here, we believe that if we introduce the possibilities of duplication in research projects, less responsive market entry by second-stage researchers, finite elasticity of supply of research, or finite elasticity of demand for the final product, or economically responsive first-period research, the model would yield similar qualitative conclusions. Introduction of a lag between research effort and results would lengthen optimal patent life in some cases, but would not otherwise change our findings on the robustness of the patent instrument under competition and the general superiority of finite over infinite patent life, for innovations that are themselves essential research inputs.

References

Arrow, K.J. and Chang, S. “Optimal Pricing, Use, and Exploration of Uncertain Natural Resource Stocks.” *Journal of Environmental Economics and Management*, Vol. 9 (1982), pp. 1-10.

Barzel, Y. “Optimal Timing of Innovations.” *Review of Economics and Statistics*, Vol. 50 (1968), pp. 348-355.

Brooks, R., Murray, M., Salant, S. and Weise, J.C. “When is the Standard Analysis of Common Property Extraction Under Free Access Correct? A Game-Theoretic Justification for Non-Game-Theoretic Analysis.” *Journal of Political Economy*, Vol. 107 (1999), pp.843-858.

Chang, H. “Patent Scope, Antitrust Policy, and Cumulative Innovation.” *RAND Journal of Economics*, Vol. 26 (1995), pp. 34-57.

Denicolo, V. “Two-stage Patent Races and Patent Policy.” *RAND Journal of economics*, Vol. 31 (2000), pp. 488-501.

Gallini, N.T. "Patent Policy and Costly Imitation." *RAND Journal of Economics*, Vol. 23 (1992), pp. 52-63.

Gilbert, Richard J. and Carl Shapiro. “Optimal Patent Length and Breadth.” *RAND Journal of Economics*, Vol. 21 (1990), pp. 106-112.

Green, J. and Scotchmer, S. “On the Division of Profit in Sequential Innovation.” *RAND Journal of Economics*, Vol. 26 (1995), pp. 20-33.

Heller, M.A. and Eisenberg, R.S. “Can Patents Deter Innovation? The Anticommons in Biomedical Research.” *Science*, Vol. 280 (1998), pp. 698-701.

Hopenhayn, H.A. and Mitchell, M.F. “Innovation Variety and Patent Breadth.” *RAND Journal of Economics*, Vol. 32 (2001), pp. 152-166.

Kitch, E.W. “The Nature and Function of the Patent System.” *Journal of Law and Economics*, Vol. 20 (1977), pp. 265-290.

Klemperer, Paul. “How Broad Should the Scope of Patent Protection be?” *RAND Journal of Economics*, Vol. 21 (1990), pp.113-130.

Koo, Bonwoo. “The Economics of Plant Genetic Resources: The Effects of Alternative Intellectual Property Protection Systems and Advances in Biotechnology” Unpublished Ph.D. dissertation, Department of Agricultural and Resource Economics, University of California, Berkeley, December, 1998.

Lee, T. and Wilde, L. “Market Structure and Innovation: A Reformulation.” *Quarterly Journal of Economics*, Vol. 94 (1980), pp. 429-436.

Loury, G. “Market Structure and Innovation.” *Quarterly Journal of Economics*, Vol. 93 (1979), pp. 395-410.

Merges, R.P., and Nelson, R. “On the Complex Economics of Patent Scope.” *Columbia Law Review*, Vol. 90 (1990), pp. 839-916.

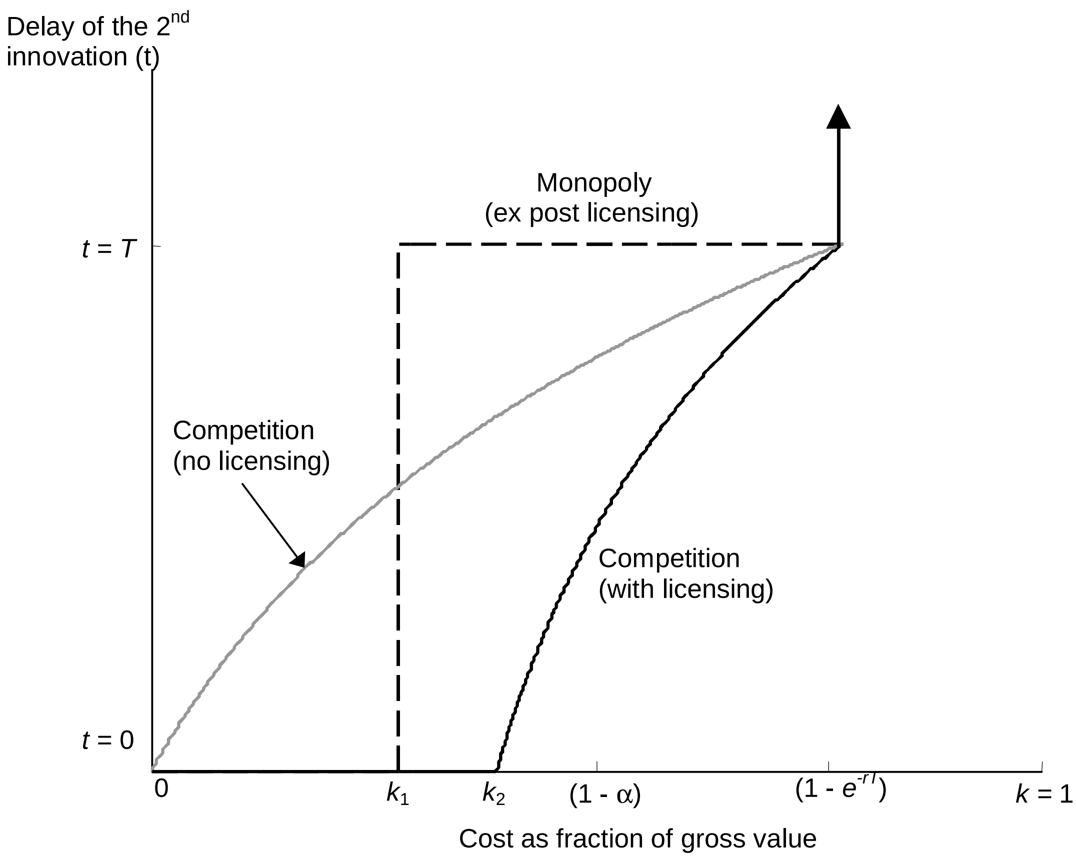
Nordhaus, W. *Innovation, Growth, and Welfare*. Cambridge, Mass.: MIT Press, 1969.

O’ Donoghue, T. “A Patentability Requirement for Sequential Innovations.” *RAND Journal of Economics*, Vol. 29 (1998), pp. 654-679.

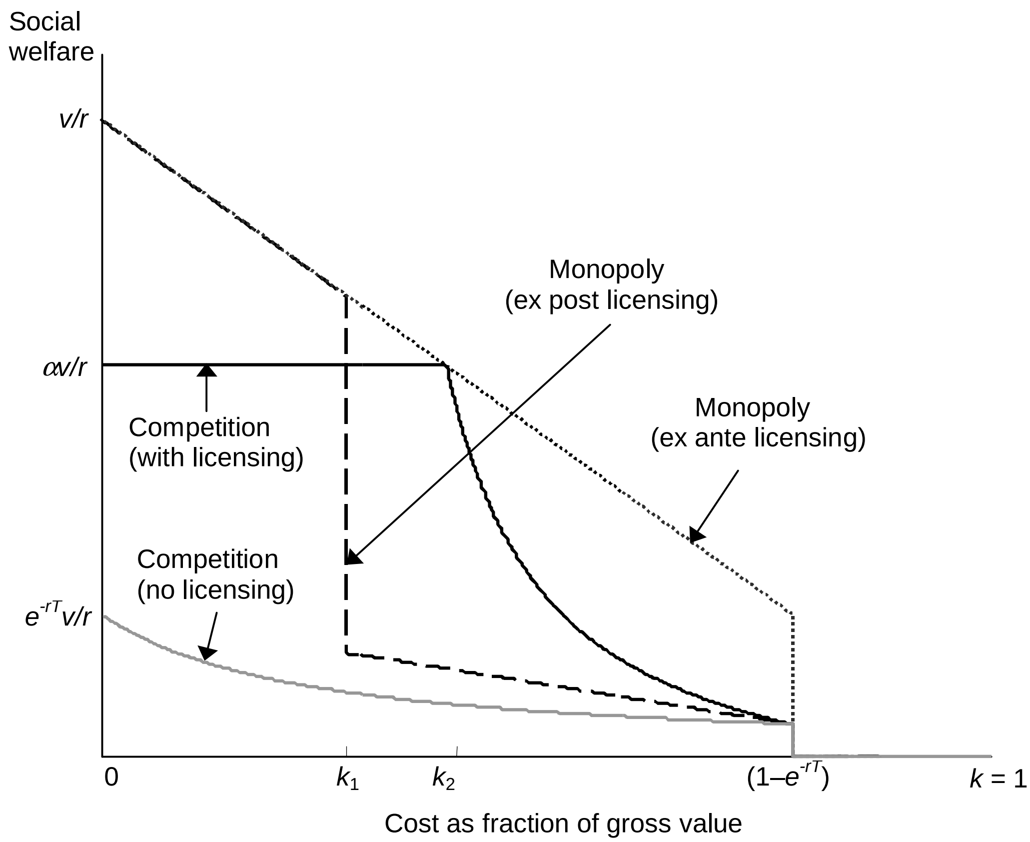
O’ Donoghue, T., Scotchmer, S. and Thisse, J. “Patent Breadth, Patent Life, and the Pace of Technological Progress.” *Journal of Economics and Management Strategy*, Vol. 7 (1998), pp. 1-32.

Wright, B.D. “The Resource Allocation Problem in R&D.” In G. Tolley, J. Hodge, and J. Oehmke, eds., *The Economics of R&D Policy*. New York: Praeger Press, 1984, pp. 41-69.

[Figure 1] Timing of the second innovation under fixed patent life

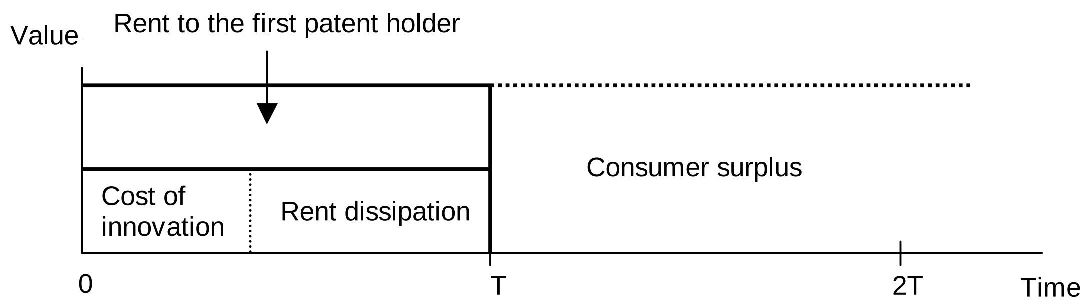


[Figure 2] Social welfare for different market structures under fixed patent life.

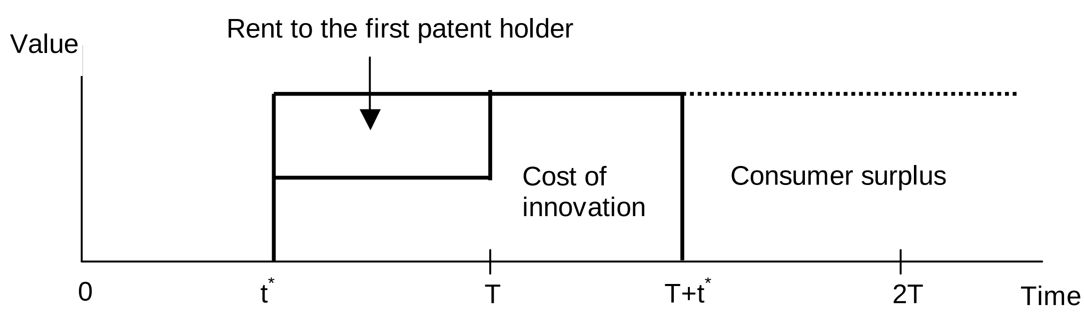


[Figure 3] Timing of innovation and social welfare under competition

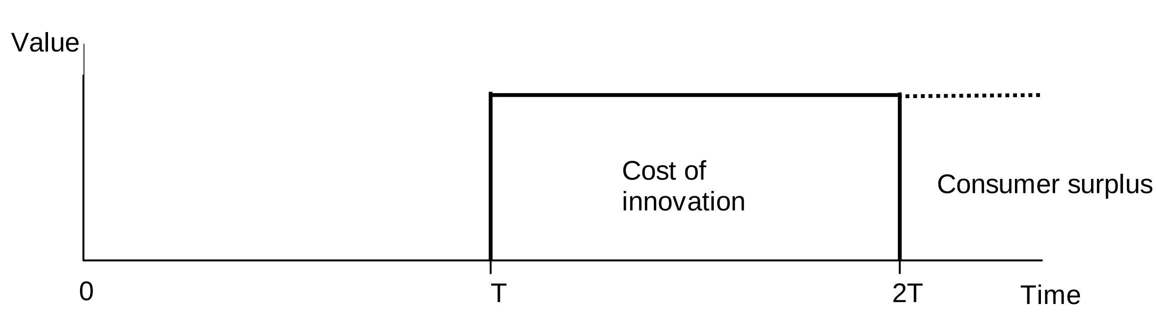
Case a: Immediate innovation at $t = 0$ (when $k < k_1$)



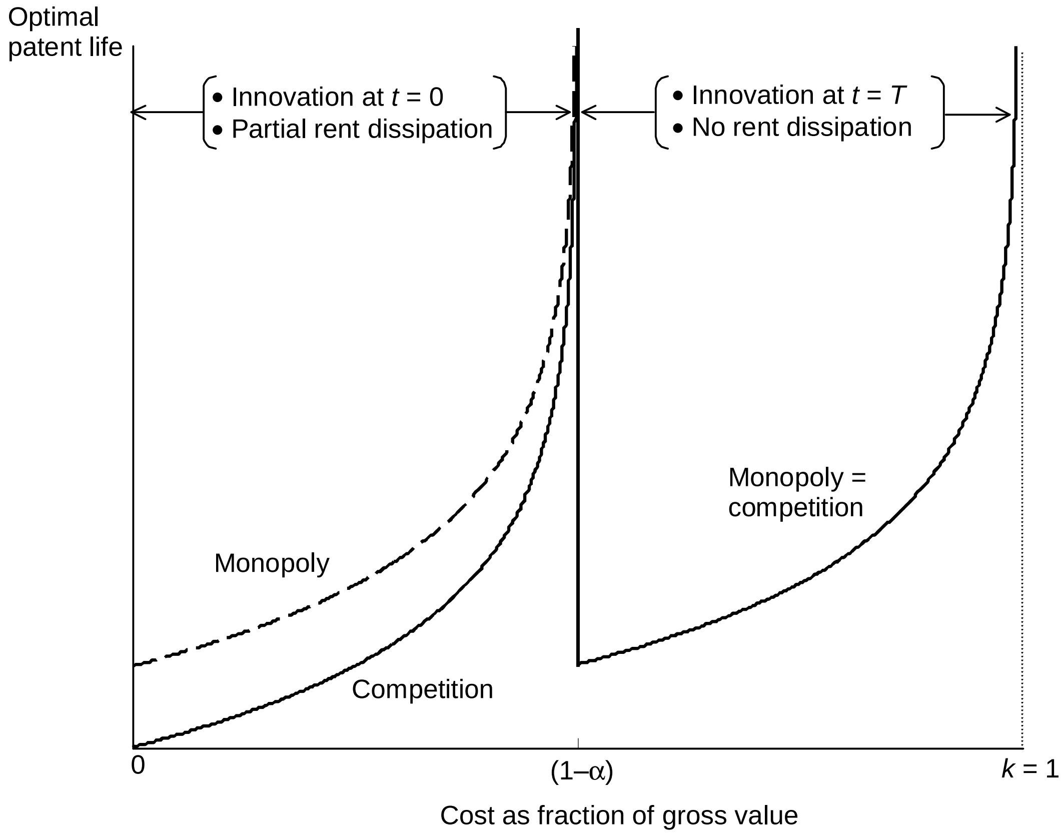
Case b: Delayed innovation at $t = t^* < T$ (when $k_1 < k < k_3$)



Case c: Innovation at $t = T$ (when $k = k_3$)



[Figure 4] The effect of innovation cost on optimal patent life



Note: There is no licensing when patent life is optimized.

[Figure 5] Sensitivity of social welfare under alternative patent lives