



**UCD GEARY INSTITUTE
DISCUSSION PAPER
SERIES**

Optimal Patent Length

Professor James Bergin

UCD Geary Institute & UCD School of Economics,
University College Dublin, Belfield, Dublin 4
Tel: 00353 1 7164618
berginj@ucd.ie

Version: December 2007

18th March 2008

The views expressed here do not necessarily reflect those of the Geary Institute.
All errors and omissions remain those of the author.

Geary WP/8/2008

Optimal Patent Length

by

James Bergin

Geary Institute School of Economics

UCD Belfield, Dublin 4

Version: December 2007

Abstract

The intent of the patent system is to encourage innovation by granting the innovator exclusive rights to a discovery for a limited period of time: with monopoly power, the innovator can recover the costs of creating the innovation which otherwise might not have existed. And, over time, the resulting innovation makes everyone better off. This presumption of improved social welfare is considered here. The paper examines the impact of patents on welfare in an environment where there are large numbers of (small) innovators — such as the software industry. With patents, because there is monopoly for a limited time the outcome is necessarily not socially optimal, although social welfare may be higher than in the no-patent state. Patent acquisition and ownership creates two opposing incentives at the same time: the incentive to acquire monopoly rights conferred by the patent spurs innovation, but subsequent ownership of those rights inhibits innovation (both own innovation and that of others). On balance, which effect will dominate? In the framework of this paper separate circumstances are identified under which patents are either beneficial or detrimental to innovation and welfare; and comparisons are drawn with the socially optimal level of investment in innovation.

1 Introduction.

Patents have existed since the late Middle Ages. In England, Edward III granted a patent for Woolen weaving in 1331 to John Kempe of Flanders. Henry VI granted a patent for the manufacture of colored glass in 1449. In this period, one major purpose in issuing patents was to stimulate growth of new manufacturing,¹ but the potential for patents to encourage innovation was also understood. The Venetian Senate voted a patent law governing all classes of invention into existence in 1474, giving patent protection for 10 years, with free access to the government. According to the preamble to the law [30]: “We have among us men of great genius, apt to invent and discover ingenious devices Now, if provisions were made for the works and devices discovered by such persons, so that others who may see them could not build them and take the inventors honor [sic] away, more men would then apply their genius, would discover, and would build devices of great utility to our commonwealth.”

Thus, by the end of the Middle Ages, at least, the use of patents to encourage the creation of new inventions and discoveries and to promote general welfare was recognized. In the New World, patents were granted by colonial governments. Massachusetts granted a patent for a new way of making salt in

¹Henry VI granted a stained glass making patent to a glass maker from Flanders, John of Utynam, with a view to developing glass making in England. In this case, the purpose of the patent was to encourage local development of a known procedure.

1641. That year the colonial legislature of Massachusetts enacted a law asserting that no monopolies (exclusive rights) would be granted [1], with the exception of new inventions “profitable to the country, and that for a short time.” The first American patent was granted in 1790, and the first French patent in 1791. By the late 1800’s, most European countries had patent laws in place. The importance attached to patents in the United States can be seen from the fact that the first ones issued were signed by George Washington, Thomas Jefferson and Edmund Randolph (the first Attorney General). Article 1, section 8 of the United States Constitution (ratified in 1788) gave Congress the power “To promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries.” After some chaos in developing procedures, the Patent Act of 1836 established the Patent Office as a separate bureau of the State Department, later (1926) to become a bureau of the Commerce Department.²

One widely noted reason for patent protection is that research costs may be recouped, hence encouraging research and development. A leading example of this kind is drug development where considerable investment outlay is required (for example to conduct clinical trials). The patent system also provides a means for independent inventors to participate in discovery and innovation. But, despite the long history of patents, the efficacy of patents in promoting innovation is still intensely debated. Because a patent confers a monopoly right, concern with the granting of patents has existed from the earliest times. Rulers granted patents as a (cost-free) means of payment for service, sometimes granting patents for procedures such as the making of salt or soap where the method of manufacture was already well known. In response to dissatisfaction with the patent system, Queen Elizabeth I issued a proclamation in 1601, allowing any individual to challenge a patent in court.³ These concerns were also present in the United States. At the time of drafting the constitution, Jefferson expressed concern to Madison in a letter dated July 31, 1788 [28]:

“..... The saying there shall be no monopolies lessens the incitements to ingenuity, which is spurred on by the hope of a monopoly for a limited time, as of 14 years; but the benefit even of limited monopolies is too doubtful to be opposed to that of their general suppression.”

So, for almost as long as they have existed, careful reflection led to an equivocal view of patents involving an inseparable mixture of good and bad — with the prevailing belief that patents sped up the rate of innovation, and the benefits from this outweighed the monopoly cost associated with the creation of temporary monopoly.

In those times, the concern was explicitly with monopoly power. While the problem of monopoly is well understood, there are many ways in which the assignment of monopoly rights through a patent generate unforeseen consequences.⁴ The literature on patents is extensive and spans not only economic issues but also the role and operation of the legal system and the patent office which generates and administers patents. To a substantial extent, the functioning of the patent system depends on the

²See [33] for a broad review of the history and patent literature to the present. A primer on the economics of patents is given in [23] and more detailed discussion in [33].

³See, for example [7], for a discussion of patent history and such issues.

⁴For example, strategic use of patents to preempt or “hold-up” a competitor or area of development frequently occurs [7]. Similarly, strategies such as the threat of costly litigation may be used to extract unwarranted rents.

efficacy of these institutions to allocate rights and incentives correctly. Many important issues relate to these institutions (such as the general quality of patents issued by the patent office and the ability of the courts to assess the legitimacy of patent claims) and will not be discussed here. (See, for example, [20], [21], and [22].) In this paper, primary focus is on the fundamental question of the tradeoff between spurred innovation and monopoly costs, leaving aside the many other issues relating to incentives, proper assignment of rights, and so on. In this idealized context one can consider and evaluate the role and functioning of the patent system on a theoretical basis; focusing on the question of whether and to what extent the system provides encouragement to innovate — the central motivation for issuing patents.

Conventional wisdom on the role of the patent system is that the granting of exclusive use of discovery for a limited period encourages innovation, and the value of this in the long run outweighs the inefficiencies associated with temporary monopoly power over the discovery. Yet, many consider this a largely unproven belief. One early and different view of incentives in this context sees competition as the key force behind innovation [15]: innovators gain market share, so the need to survive places continuous pressure on firms to innovate. In this Schumpeterian view, it is the absence of protection that drives firms to innovate. In the area of medical research, see [9], [14], and the references cited therein for a mixed assessment of the value of patents. (See [8] for a recent and broad ranging critique of the patent system, and the notion of intellectual property more generally.) Concerning software patents, Bessen and Hunt [11] conclude that in the software industry there appears to be little correlation between the rate at which firms invest in R&D and the rate of innovation. The impact of patents may depend substantially on the field of application: Bessen and Hunt identify the chemical industry as one where patents may be important in the decision to conduct R&D. Thus the balance of costs and benefits is still the subject of intense debate; and the intent of this paper is to address this matter, comparing the benefit from spurred invention against the temporary monopoly costs incurred.

While the use of (voluntary) licencing potentially permits broad application of new discovery immediately, this depends on the feasibility of such agreements. And, one may appeal to the Coase theorem to assert that where gains are possible, licencing agreements will be reached. However, in practice, with a multiplicity of “players”, unclear ownership rights or breadth of patent, the potential use of hold-up tactics to extract rents, the complexity, time and expense of legal resolution, and so forth, the expectation of a Coase-style resolution may be wishful thinking. The study of such issues will not be taken up here. This paper considers the question of whether the welfare cost from temporary monopoly is greater or less than the welfare gains from increased innovation spurred by the monopoly rights granted by patents. And, in examining this tradeoff, the research here is concerned with the impact of patents in a specific class of environments with features similar to that of the software industry.

In the specific case of software development there are a number of notable features. Development is typically in small incremental steps involving the combination or modification and extension of existing ideas. Work is highly correlated across developers because, for example, a program will typically make use of many different techniques so that overlap is natural. Furthermore, development proceeds in a sequential and incremental manner, typically at a rapid pace. And, because entry costs to writing software are very low,⁵ there are very many individuals and companies working in the field. What is the

⁵Although, in certain circumstances, such as the development of operating systems that scale to large multiprocessor

impact of patenting on societal welfare and on the rate of innovation in an industry with these features? In contrast to some areas of innovation (such as drug or microchip development, for example) where patents are prevalent, in the area of software capital costs are relatively low and, as mentioned, development can be highly correlated, even contemporaneous.⁶ The impact of patents in software development has been studied elsewhere: Bessen and Maskin [12] consider environments where innovation is sequential and complementary — successive innovation builds on what has gone before in a sequential way and innovation is complementary in the sense that the probability of success in discovery is improved when more firms pursue research. Different innovators follow different routes of research and in this setting they show how patents can actually inhibit innovation by limiting imitation that spurs the development of further innovation. They focus on the case where there are a small number of competing innovators. In contrast, here the environment is one with a large population.

An outline of the paper is as follows. Section 2 describes the environment and provides some remarks on the recent history of software patenting. Section 3 develops a model for this and similar industries. The welfare implications of the patent system are considered in detail in section 4. Section 5 considers the consequences of changing patent length and highlights the impact of access to the innovation of others on the marginal product of investment and the manner in which such knowledge correlates with a firms' own technology in terms of payoff impact. Depending on these features, patents may either reduce or raise social welfare (although they never lead to the socially optimal level of innovation). Section 6 concludes.

2 The Environment.

In certain notable areas such as drug development considerable investment outlay is required (for example to conduct clinical trials). This appears as an argument in favor of patent protection: research costs may be recouped, hence encouraging research and development. In the class of environment discussed here, notably the software environment, development follows a very different pattern. Development is typically in small incremental steps and proceeds in a sequential manner. Work is highly correlated across developers: for example, a program will routinely make use of many different techniques so that overlap is natural. Development routinely involves the combination or modification and extension of existing ideas and computer code and the pace of development is rapid. Entry costs to the software industry are usually very low, so there are very many individuals and companies working in the field, from small one-person companies, to large corporations.⁷

computers, significant hardware investment and related support may be important.

⁶In such environments, it becomes more difficult to unravel competing ownership rights and so the legal rights of the patentee may be clouded, or there may be multiple entities with rights to various components of a program or implementation of an idea. In such circumstances, the detailed operation of the patent office and the court system is central to any attempt to disentangle rights and claims. However, this paper is not concerned with such important matters that require separate consideration. The intent here is to explore the basic question of the social value of patents.

⁷The current linux kernel under development (2007, version 2.6.22) has over 900 developers, averages 4 changes per hour, has over 8 million lines of code and runs on 1 to 4096 processor computers.

2.1 Software and patents.

Until the early 1980's software was not patentable, but around this time (in 1983), the USPTO gradually extended the notion of patentability to include equipment whose only novel feature was the use of computer software to manage the equipment. One of the earliest software patents was issued for a rubber curing process, whereby a program monitored the temperature during the curing process.⁸ Over time, software became patentable in the United States in standalone form and now almost any software application is patentable. Business methods in particular have received attention with such high profile patents as the "One-Click" patent of Amazon. Mathematical procedures have also been patented. Patent 6,434,582 provides an algorithm for computing the cosine of a "relatively small angle"; patent 6,078,938 provides a procedure for solving systems of linear equations. This contrasts with expectation. In the mid 1970's the National Commission on New Technological Uses of Copyrighted Works wrote, regarding patents: "Even if patents prove available in the United States, only the very few programs which survive the rigorous application and appeals procedure could be patented".⁹ In the twenty-two year period since the USPTO began issuing software patents, 150,000 patents were issued. In contrast, the (long established) pharmaceuticals industry received 80,000 patents in that period [27].¹⁰ Given this rapid expansion in the number of patents granted, it is natural to revisit the question of how the system affects the pace of investment and innovation, and examine the overall impact on societal welfare.

3 The Model.

The model considers a single market supplied by many firms. Firms are differentiated by their (cost) efficiency which depends on own technology and the prevailing state of the art that is publicly available for use. We assume that firms protect own technology improvements through patenting, so that only the technology of a firm that is beyond the patent life is available for use by competitors. Demand may depend on the state of technology, so that technological improvements can push demand up over time.

The discussion abstracts from some important issues: it is assumed that only "genuine" innovations are patented, and all such innovations are patented. Patent holders do not licence innovations but exploit them as monopolists until patent expiry. (See section 5.3.1 for some additional comments.) In this framework, the merits of patenting stand or fall on whether or not monopoly for a limited period of time has the overall effect of encouraging innovation, relative to the no-patent environment. Because the market does not internalize fully the benefits of innovation there is scope for welfare improvement through incentives to encourage innovation — a point briefly illustrated in section 5.3.1. However, the study of welfare improvement by particular policy schemes is not considered here.

⁸Essentially, the program implemented the Arrhenius equation — an equation relating the rate of a chemical reaction to the temperature. $k = Ae^{(-\frac{E_a}{R \times T})}$ where k is the rate coefficient, A is a constant, E_a is the activation energy, R is the universal gas constant, and T is the temperature (in degrees Kelvin).

⁹*Final Report of the National Commission on New Technological Uses of Copyrighted Works*, 1978, chapter 3: Computers and Copyright, p17.

¹⁰It is widely believed that the standard for obtaining a software patent is low [20], so that a large number of patents issued could not survive close scrutiny.

3.1 The Model: Main Features

The technology of each firm evolves over time. At any point in time, a firm has a fixed technology, has access to technologies that are publicly available, and makes investment decisions that affect its future technology. Aggregating individual behavior gives the aggregate distribution over technology and investment and determines the evolution of the distribution on technology over time. These details are described next.

3.1.1 Technology.

An enterprise is characterized by its technology $\alpha \in \mathcal{A}$, where \mathcal{A} is an ordered space, with order \succeq .¹¹ This formulation permits a large set of technologies and, in particular, allows different firms to have different strengths and weaknesses. The distribution of technologies in the market is denoted μ , or μ_t to denote the distribution of technologies at time t : a probability measure on \mathcal{A} . If α and α' are in the support of μ_t , they represent two technologies in operation at time t . They are comparable if $\alpha \succeq \alpha'$ in which case α is a better firm than α' in every way; but in general technologies may not be comparable (neither $\alpha \succeq \alpha'$ or $\alpha' \succeq \alpha$). Let

$$\boldsymbol{\mu}_t = \{\mu_\tau\}_{\tau=t}^{-\infty} = (\mu_t, \dots, \mu_0, \mu_{-1}, \dots) = (\mu_t, \dots, \mu_{t-\ell+1}; \mu_{t-\ell}, \dots, \mu_0, \mu_{-1}, \dots)$$

denote the sequence of past technology distributions over time. With patent length ℓ , technologies $\{\mu_{t-\ell-j}\}_{j \geq 0}$ are available for public use at time t : any technology in the support of $\mu_{t-\ell-j}$ may be used by a firm. At period t , technologies older than ℓ are available for use by all firms: apart from technology α , only technologies in the support of distribution $\mu_{t-\ell}$ or older may be used by firm α . All innovations are patent protected but not licensed.

3.1.2 The Firm

Firms supply a market with demand $P_d(Q, \boldsymbol{\mu})$ reflecting the assumption that new technology raises demand. If demand depends only on current technology, then $P_d(Q, \boldsymbol{\mu}_t) = P_d(Q, \mu_t)$, and this will be assumed.¹² Firms are distinguished by technology in two ways: own technology enters the cost function directly, and the history of aggregate technologies is also observed.

The cost function of an enterprise depends on its current technology, α , the history of technologies, $\boldsymbol{\mu}$, and patent length: $c(q, \alpha, \boldsymbol{\mu}, \ell)$. Assume that c is weakly decreasing in α and weakly increasing in ℓ : better private or public technology can only lower cost.¹³ In this notation, the argument ℓ identifies

¹¹ \mathcal{A} is an ordered topological space where the relation \succeq is reflexive ($\alpha \succeq \alpha$); transitive ($\alpha \succeq \alpha'$ and $\alpha' \succeq \alpha''$ imply $\alpha \succeq \alpha''$); and antisymmetric ($\alpha \succeq \alpha'$ and $\alpha' \succeq \alpha$ imply $\alpha = \alpha'$). (For example: $\mathcal{A} = \{\alpha \mid \alpha : [a, b] \rightarrow \mathfrak{R}, \alpha \text{ measurable}\}$ where $\alpha' \succeq \alpha$ if $\alpha'(x) \geq \alpha(x), x \in [a, b]$.) If technology were characterized by a real number, the firm with the largest α would be the best firm, unequivocally, eliminating the possibility for different firms to have area specific strengths.

¹²It may be more realistic to consider a model where different technology vintages are viewed as separate markets with newer technologies possibly having higher demand growth. However, this would greatly complicate the welfare analysis. Here, firms with poorer technologies will suffer relatively lower profit, so that the investment incentives are similar.

¹³If the patent length, ℓ , lies between two periods, τ and $\tau - 1$, let $(\ell - [\tau - 1])\mu_\tau + (\tau - \ell)\mu_{\tau-1}$ be the most recent publicly available technology. The corresponding cost may have the form $(\ell - [\tau - 1])c(q, \alpha, \mu_\tau, \ell) + (\tau - \ell)c(q, \alpha, \mu_{\tau-1}, \ell)$.

the $(\ell + 1)^{\text{th}}$ element of the list $\boldsymbol{\mu}$ as publicly usable. Since a firm cannot use patented technology, $c(q, \alpha, \boldsymbol{\mu}_t, \ell)$ does not depend on $(\mu_t, \dots, \mu_{t-\ell+1})$. In particular, a variation in aggregate behavior at time t (affecting the aggregate distribution in subsequent periods), has no impact on cost in period $t + j$ if $j < \ell$ — since the aggregate distribution at dates closer to t than $t - \ell$ do not affect the cost at time t . And, if the distribution of technologies improves every period, any technology present at time $t - \ell - 1$ will be dominated by a technology present at period $t - \ell$. In these circumstances, the cost function has the form $c(q, \alpha, \mu_{t-\ell})$. This will be assumed throughout, but to maintain notational consistency with usage elsewhere, the notation $c(q, \alpha, \boldsymbol{\mu}, \ell)$ will apply — with the understanding that the ℓ^{th} element of $\boldsymbol{\mu}$ affects cost. Finally, assume that better publicly available technology can only improve cost: if $\mu'_{t-\ell} \succcurlyeq \mu_{t-\ell}$ ¹⁴ then $c(q, \alpha, \mu'_{t-\ell}) \leq c(q, \alpha, \mu_{t-\ell})$. In terms of $\boldsymbol{\mu}$ notation, given $\boldsymbol{\mu}_t = (\mu_t, \mu_{t-1}, \dots)$ and $\boldsymbol{\mu}'_t = (\mu'_t, \mu'_{t-1}, \dots)$, if $\mu'_{t-\ell} \succcurlyeq \mu_{t-\ell}$, then $c(q, \alpha, \boldsymbol{\mu}'_t, \ell) \leq c(q, \alpha, \boldsymbol{\mu}_t, \ell)$.

At output q and price p profit is:

$$\pi(p, q, \boldsymbol{\mu}, \alpha, \ell) = pq - c(q, \boldsymbol{\mu}, \alpha, \ell),$$

so that profit maximization gives¹⁵

$$p - c_q(q, \boldsymbol{\mu}, \alpha, \ell) = 0,$$

with solution $q(p, \boldsymbol{\mu}, \alpha, \ell)$ and corresponding profit, $\pi(p, \boldsymbol{\mu}, \alpha, \ell)$. Given p , these functions depend only on $\mu_{t-\ell}$ through $\boldsymbol{\mu}_t$. At price p , total supply is $Q_s(p, \boldsymbol{\mu}, \ell) = \int q(p, \boldsymbol{\mu}, \alpha, \ell) \mu(d\alpha)$ when the current aggregate distribution is μ . Let $P_s(Q, \boldsymbol{\mu}, \ell)$ be the inverse supply function. Market clearing gives $P_s(Q^*, \boldsymbol{\mu}, \ell) = P_d(Q^*, \boldsymbol{\mu})$, with market clearing price $p^* = p^*(\boldsymbol{\mu}, \ell)$ and quantity $Q^* = Q^*(\boldsymbol{\mu}, \ell)$. Equilibrium profit of firm α may be written as $\pi(p^*, \boldsymbol{\mu}, \alpha, \ell)$ or as $\pi(\boldsymbol{\mu}, \alpha, \ell)$. Finally, over time firms invest to improve future technology: the level of investment i costs $r(i)$, where r is assumed to satisfy $r' \geq 0$ and $r'' \geq 0$.

3.1.3 The Evolution of Technology

A firm may invest from period to period in research and technological improvement. Technological improvement is represented by a transition kernel, $P(d\tilde{\alpha} \mid \boldsymbol{\mu}, \alpha, i, \ell)$, where i is investment. As with earlier notation, this may alternatively be written as $P(d\tilde{\alpha} \mid \mu_{t-\ell}, \alpha, i)$: a firm with technology α may use investment and the knowledge of technologies in the support of $\mu_{t-\ell}$ to develop its technology next period. This formulation allows for the possibility that a firm may require investment to achieve the standards represented by $\mu_{t-\ell}$ — it may not be possible for a firm to effortlessly implement the best technologies in the support of $\mu_{t-\ell}$.

The key assumptions on the transition kernel are that better firms are more likely to draw better technology, more investment improves the chances of drawing a good technology, and less access to

¹⁴Given two measures $\mu, \nu \in \mathcal{P}(\Lambda)$, μ dominates ν , written $\mu \succcurlyeq \nu$ if and only if for all measurable increasing functions $g : \Lambda \rightarrow \mathfrak{R}$, $\int g d\mu \geq \int g d\nu$. (Note that “ \succeq ” is an ordering on Λ , and “ \succcurlyeq ” an ordering on $\mathcal{P}(\Lambda)$). See Appendix III for further discussion.

¹⁵Assume that marginal cost is increasing in q to ensure that the first order condition gives a maximum. Furthermore, assume that both cost, c and the transition kernel, P , are continuous functions of all arguments.

technology of others worsens the chances of drawing a good technology (T1). Furthermore, technology can only improve over time (T2), and better technology firms are more successful innovators (T3). Formally:

- T1. $P(d\tilde{\alpha} \mid \boldsymbol{\mu}, \alpha, i, \ell)$ is weakly increasing in α , and in i ; and weakly decreasing in ℓ .¹⁶ Furthermore, $P(d\tilde{\alpha} \mid \boldsymbol{\mu}, \alpha, i, \ell)$ is increasing in $\boldsymbol{\mu}$ — in the sense that if $\boldsymbol{\mu}'$ dominates $\boldsymbol{\mu}$ coordinate-wise, then other things equal, a better distribution is drawn conditional on $\boldsymbol{\mu}'$ than $\boldsymbol{\mu}$. Better technology or higher investment improve next periods distribution; longer patent life reduces access to technology so can only worsen next periods distribution.
- T2. If $\alpha' \in \text{supp } P(\cdot \mid \boldsymbol{\mu}, \alpha, i, \ell)$, then $\alpha' \succeq \alpha$, where given a measure ν on Λ , $\text{supp } \nu$ is the support of ν .
- T3. The marginal productivity of investment weakly increases with α . For any continuous increasing monotone function g on Λ :¹⁷

$$\alpha' \succeq \alpha \quad \text{implies} \quad \int g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} \mid \boldsymbol{\mu}_t, \alpha', i, \ell) \geq \int g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} \mid \boldsymbol{\mu}_t, \alpha, i, \ell). \quad (1)$$

The investment strategies of firms in conjunction with the transition kernel, P , move the state of the system forward over time. This is discussed in the next section.

REMARK 1. Thus, improvement in technology overall results from the flow of individual discoveries — with each individual discovery insignificant relative to the overall volume of discovery. One possible extension of this model is to allow for “paradigm shift” discoveries which revolutionize an industry — in the way, for example, that the *Mosaic* web browser revolutionized internet use. The formulation used here can accommodate such an extension provided that big breakthroughs are unanticipated. In such a formulation, there is positive probability of a breakthrough discovery in any period (some firm will have a major discovery or development), but no single firm can guarantee that it will have such a discovery with positive probability. In this case, revolutionary innovations are unanticipated and hence don't affect the investment incentives of firms.

3.1.4 Firms Strategies and the Evolution of Technology.

Firms strategies are represented by a joint distribution, τ , on $(i, \alpha) \in I \times \mathcal{A}$, written $\tau \in \mathcal{M}(I \times \mathcal{A})$. Conditioning on α , $\tau(di \mid \alpha)$, gives the distribution over investment of firm α . Given the extant distribution over technologies is μ , for consistency, if $\tau \in \mathcal{M}(I \times \mathcal{A})$ the marginal distribution of τ on \mathcal{A} should coincide with μ : $\text{marg}_{\mathcal{A}} \tau = \mu$. Let $\mathcal{C}(\mu) = \{\tau \mid \text{marg}_{\mathcal{A}} \tau = \mu\}$, the set of distributions on $I \times \mathcal{A}$ with marginal μ on \mathcal{A} . The distribution of technologies evolves as:

$$\mu_{t+1}(\cdot) = \int_{i_t, \alpha_t} P(\cdot \mid \boldsymbol{\mu}_t, \alpha_t, i_t, \ell) \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) = \int_{i_t, \alpha_t} P(\cdot \mid \boldsymbol{\mu}_t, \alpha_t, i_t, \ell) \tau_t(di_t \times d\alpha_t), \quad (2)$$

¹⁶For example, $\alpha' \succeq \alpha$ implies that $P(d\tilde{\alpha} \mid \boldsymbol{\mu}, \alpha', i, \ell) \succcurlyeq P(d\tilde{\alpha} \mid \boldsymbol{\mu}, \alpha, i, \ell)$.

¹⁷Assume that $\frac{1}{i' - i} [P(d\tilde{\alpha} \mid \boldsymbol{\mu}_t, \alpha, i', \ell) - P(d\tilde{\alpha} \mid \boldsymbol{\mu}_t, \alpha, i, \ell)]$ converges weakly to a signed measure $\Delta_i P(d\tilde{\alpha} \mid \boldsymbol{\mu}_t, \alpha, i, \ell)$ as $i' \rightarrow i$.

So, given the current distribution on technologies, μ_t , if α_t invests according to the strategy $\tau_t(\cdot \mid \alpha_t)$, then next period the aggregate distribution on technologies is given by μ_{t+1} . The distribution $\mu_{t+1}(\cdot)$ depends on μ_t , ℓ , τ_t and μ_t . This may be made explicit by writing:

$$\mu_{t+1}(\cdot) = \varphi_{t,1}(\cdot \mid \mu_t, \tau_t, \ell), \quad \tau_t \in \mathcal{M}(I \times \mathcal{A}) \quad (3)$$

where $\tau_t \in \mathcal{C}(\mu_t)$. Appendix I describes in detail the evolution of the individual and aggregate distributions over time.

4 Welfare and Efficiency.

In this environment, welfare is most naturally evaluated by the sum of consumer and producer surplus, and in the multi-period context by the present value of the surplus flow. Letting $\{p_t^*\}$ be the sequence of market clearing prices, the present value of consumer surplus is given by $\sum_{t=1}^{\infty} \delta^{t-1} [P_d(Q_t, \mu_t) - p_t^*] = PV_{cs}$. Profit net of investment cost for firm α_t is $[\pi(\mu_t, \alpha_t, \ell) - r(i_t)]$ and the present value, aggregating over all firms is $PV_{ps} = \sum_{t=1}^{\infty} \delta^{t-1} \int [\pi(\mu_t, \alpha_t, \ell) - r(i_t)] \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t)$. Together, these equal the discounted sum of the areas between the supply and demand curves. So, the welfare at time t generated in the market in equilibrium, as measured by total surplus, is given by:

$$PV_{ts}(\mu_t, \ell) = PV_{cs} + PV_{ps}. \quad (4)$$

Equivalently, this is measured by the present value of the area between demand and supply each period, less investment. Let

$$S(\mu, \ell) = \max_Q \int_0^Q [P_d(Q, \mu) - P_s(Q, \mu, \ell)] dQ, \quad (5)$$

and define a Bellman equation on surplus:

$$V(\mu, \ell) = \max_{\tau \in \mathcal{C}(\mu)} \{S(\mu, \ell) - \int r(i) d\tau + \delta V(\mu', \ell)\} \quad (6)$$

where $\mu' = (\mu', \mu)$, with μ' determined according to equation (2): $\mu'(\cdot) = \int P(\cdot \mid \alpha, \mu, i, \ell) \tau(di \times d\alpha)$. Equation (6) gives the surplus generated under optimal choice of aggregate investment each period, where the optimizing τ is the current period investment strategy across all firms. Define

$$\mathcal{C}^*(\mu) = \{\tau = (\tau_1, \tau_2, \dots) \mid \text{marg}_{\Lambda} \tau_t = \mu_t, \mu_1 = \mu, \mu_{t+1} = \int P(\cdot \mid \mu_t, \alpha_t, i_t, \ell) \tau_t(di_t \times d\alpha_t), t \geq 1\} \quad (7)$$

giving those sequences (τ_1, τ_2, \dots) of feasible strategies — consistent with the initial aggregate distribution. With this notation, the value function may be written:

$$V(\mu, \ell) = \max_{\tau \in \mathcal{C}^*(\mu)} \sum_{t=1}^{\infty} \delta^{t-1} [S(\mu_t, \ell) - \int r(i) d\tau_t]$$

$$\begin{aligned}
&= \max_{\boldsymbol{\tau} \in C^*(\boldsymbol{\mu})} \sum_{t=1}^{\infty} \delta^{t-1} [S(\boldsymbol{\mu}_t(\boldsymbol{\tau}^t), \ell) - \int r(i) d\tau_t] \\
&= \max_{\boldsymbol{\tau} \in C^*(\boldsymbol{\mu})} \hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)
\end{aligned} \tag{8}$$

where $\boldsymbol{\mu}$ is defined inductively from $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ and $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots)$, according to equation (2) (and see equation 18 in the appendix), and where $\hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)$ is the present value of the surplus flow from the strategy $\boldsymbol{\tau}$ with initial distribution $\boldsymbol{\mu} = \boldsymbol{\mu}_1$. Let $PV_{soc} = V(\boldsymbol{\mu}, \ell)$, the present value of total surplus under the welfare maximizing policy.

Taking the current period to be t , one may consider the impact of varying the current strategy, τ_t . Regardless of other period strategies, optimality requires that there is no within period variation producing a positive change in the present value of surplus. In particular, consider the impact of varying τ_t in the direction $\Delta\tau_t = (\tau'_t - \tau_t)$: $\tilde{\tau}_t = \tau_t + \epsilon\Delta\tau_t = (1 - \epsilon)\tau_t + \epsilon\tau'_t$. Write $\tilde{\boldsymbol{\tau}}$ for the resulting aggregate distribution sequence. Then, if $\boldsymbol{\tau}$ maximizes (8), necessarily:

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\hat{V}(\boldsymbol{\mu}, \tilde{\boldsymbol{\tau}}, \ell) - \hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)] \leq 0 \tag{9}$$

This computation leads to theorem 1.

REMARK 2. To simplify calculations, theorem 1 considers the case where $P_d(Q, \boldsymbol{\mu})$ is independent of $\boldsymbol{\mu}$ so that only the supply curve shifts in response to distributional changes. Also, for ease of notation, write z_s for (α_s, i_s) .

In theorem 1 the impact of a variation in the aggregate distribution, τ_t is broken into three components: the effect of the resulting aggregate variation on cost and on the transition kernel, and the direct variation on profit accruing to each firm.

Theorem 1. *At the solution to the optimization problem in equation (8), for all feasible $\Delta\tau_t$:*

$$\begin{aligned}
&\delta^{t-1} \sum_{j=1}^{\infty} \delta^j \int_{z_{t+j}} \left\{ \partial_C \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell \mid \Delta\tau_t) + \partial_T \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell \mid \Delta\tau_t) \right\} \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \\
&\quad + \delta^{t-1} \cdot \sum_{j=1}^{\infty} \delta^j \int_{z_t} \left\{ \left[\bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell) - \int_{z_{t+j}} r(i_{t+j}) \tau_{t+j}(di_{t+j} \mid \alpha_{t+j}) \psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t, l) \right] \right. \\
&\quad \left. - r(i_t) \right\} \Delta\tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \leq 0 \tag{10}
\end{aligned}$$

where $\partial_C \pi_{t+j}$, $\partial_T \pi_{t+j}$ and $\bar{\pi}_{t+j}$ capture the impact of current investment on future profit in period $t+j$ arising from:

1. cost reduction ($\partial_C \pi_{t+j}$) through improvement in aggregate technology,
2. improvement in aggregate technology at the length of the patent period ($\partial_T \pi_{t+j}$), and
3. direct impact ($\bar{\pi}_{t+j}$), from improvement in technology due to the change in investment $\Delta\tau_t(di \mid \alpha)$.

(All proofs are in appendix II.)

In the optimization problem faced by the individual firm (discussed below in section 4.1), the terms in $\partial_C\pi$ and $\partial_T\pi$ in expression (10) do not appear — since these result from variations in the *aggregate* distributions. This results in a divergence between social and private interests which is discussed further there. The expression $\bar{\pi}_{t+j}$ in the third term has the form:

$$\bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell) \stackrel{\text{def}}{=} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t, \ell) \quad (11)$$

and this represents the profit gain *directly* resulting from investment variation, and the only component to appear in the individual optimization calculation.

REMARK 3. In the full computation, taking into account the dependence of P_d on $\boldsymbol{\mu}$, an additional group of terms appear since aggregate distribution shifts “grow” demand and increase welfare.

The welfare optimization considered in equation (6) or (8) determines the socially optimal investment policy for a *given* patent life ℓ . From the social planner perspective, increasing patent length creates inefficiency since it restricts the use of best available technology. The socially optimal patent length, ℓ , is zero where all knowledge is fully utilized.

Theorem 2. *The socially optimal value of ℓ is $\ell = 0$.*

The logic for the result is simple: other things equal, a reduction in patent length benefits each firm in terms of cost reduction and improvement in innovation, leading to an increase in output and an outward shift in supply (each period) raising surplus. So, the impact effect of the reduction in ℓ is to raise surplus, *prior* to optimization over the aggregate distribution, which can only raise surplus further. Thus, $PV_{soc} \stackrel{\text{def}}{=} V(\boldsymbol{\mu}, \ell)$ is maximized at $\ell = 0$. Viewing PV_{soc} as a function of patent length, $PV_{soc}(\ell)$ is plotted in figure 3.

4.1 The Divergence of Public and Private Incentives

The discussion above considers necessary conditions for optimality of the aggregate investment strategy. In those computations, in addition to the direct effects of investment change, the externality effects from the benefit of creating know-how appear in the terms $\partial_C\pi$ and $\partial_T\pi$ in equation (10). The following discussion separates these components. From the perspective of an individual firm α , the aggregate distribution over time is a parameter in the optimization problem, which involves a sequence of investment output decisions $\{i_t, q_t\}$. At each point in time, the period t output decision is chosen as in the one period model and can be eliminated from the problem, leaving investment as the sole choice variable.

Fix an aggregate strategy sequence $\boldsymbol{\tau}^t = \{\tau_s\}_{s=t}^\infty$. This determines the aggregate distribution going forward in time $(\boldsymbol{\mu}_{t+1}, \boldsymbol{\mu}_{t+2}, \dots)$. For a firm with technology α_s at time s , let the investment strategy be $\tau_s(di \mid \alpha_s)$, so that from period t , the present value of net revenue with this strategy is, at time t :

$$\pi(p_t^*, \boldsymbol{\mu}_t, \alpha_t, \ell) - r(i_t) +$$

$$\delta \int_{z_{t+1}} [\pi(p_{t+1}^*, \boldsymbol{\mu}_{t+1}, \alpha_{t+1}, \ell) - r(i_{t+1})\tau_{t+1}(di_{t+1} | \alpha_{t+1})]P(d\alpha_{t+1} | \boldsymbol{\mu}_t, \alpha_t, i_t, \ell) + \delta^2 \dots$$

Given $\boldsymbol{\tau}^t = \{\tau_s\}_{s=t}^\infty$ and suppressing aggregate variables in the notation, let the j -period ahead distribution be $\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, \ell)$ (see section 3.1.4 and appendix I), so the full expression may be written:

$$\begin{aligned} & \pi(p_t^*, \boldsymbol{\mu}_t, \alpha_t, \ell) - r(i_t) + \\ & \sum_{j=1}^{\infty} \delta^j \left\{ \int_{z_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) - r(i_{t+j})\tau_{t+j}(di_{t+j} | \alpha_{t+j}) \right\} \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, \ell). \end{aligned} \quad (12)$$

In market equilibrium, firm α maximizes this expression by choice of i (at each period).

REMARK 4. In contrast with equation (10), in the competitive equilibrium individually optimal behavior requires that for all $\Delta\tau_t$:

$$\begin{aligned} \sum_{j=1}^{\infty} \delta^j \int_{i_t} \left\{ \bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell) - \int_{z_{t+j}} r(i_{t+j})\tau_{t+j}(di_{t+j} | \alpha_{t+j})\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, \ell) \right. \\ \left. - \int_{z_t} r(i) \right\} \Delta\tau_t(di_t | \alpha_t) \leq 0 \end{aligned}$$

implying too low a level of investment, since the positive externality effect of $\Delta\tau_t$,

$$\sum_{j=1}^{\infty} \delta^j \int_{i_t} \left\{ \partial_{C+T}\pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t) \right\} \tau_t(di_t | \alpha_t) > 0,$$

is ignored.¹⁸

In general, the market equilibrium is not socially optimal.

Theorem 3. *The Social optimum coincides with the market equilibrium if and only if*

$$\partial_C \pi_{t+j} = \partial_T \pi_{t+j} = 0, \quad j \geq 1$$

From equation (10), the terms $\partial_C \pi_{t+j}$ and $\partial_T \pi_{t+j}$ represent the (positive) externalities for individual firms (from aggregate distributional improvements). The only remaining term is the direct effect on the firms profit from investment. These terms, $\partial_C \pi_{t+j}$ and $\partial_T \pi_{t+j}$, are 0 when shifts in the aggregate technology distribution have no impact on individual costs and transition probabilities. For $j < \ell$ this is always the case since new innovations of others cannot be used; and 0 for $j > \ell$ when the evolution of technology is fast — so that what is in the public domain is completely outdated. Thus, social optimality coincides with the competitive outcome only when there are no externality benefits in either cost reduction or technological development. (The proof in appendix II formalizes these remarks.)

¹⁸Where, for ease of notation:

$$\partial_{C+T}\pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t) = \partial_C \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t) + \partial_T \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t)$$

With externalities associated with patented information, at $\ell = 0$, $\partial_C \pi_{t+j} \neq 0$ and $\partial_T \pi_{t+j} \neq 0$ for some $j \geq 1$, in which case the social optimum differs from the market equilibrium outcome. However, when ℓ is large, information is old before it comes into the public domain for public use. In environments where there is rapid technological progress, old information is useless — in the sense that the technologies of that vintage are superseded by the current technology of any firm.

With a patent life of ℓ periods, the distribution of firm α_t 's technology at time $t + 1$ given by

$$P(\cdot \mid (\mu_t, \dots, \mu_{t-\ell+1} : \mu_{t-\ell}, \dots), \alpha_t, i, \ell),$$

so that at time t , all technologies in existence at time $t - \ell$ are publicly available as inputs in research and development.

Definition 1. *Say that the pace of innovation is fast relative to patent life if*

$$\alpha_t \in \text{supp } \mu_t \text{ and } \tilde{\alpha} \in \text{supp } P(\cdot \mid \mu_t, \alpha_t, i_t, \ell) \text{ implies that } \tilde{\alpha} \succeq \alpha', \forall \alpha' \in \text{supp } \mu_{t-\ell+1}.$$

If the pace of evolution of technology is fast, then the technology drawn for time $t + 1$ (based on own technology α_t and publicly available information $\mu_{t-\ell}$) dominates technologies that become publicly available at time $t - \ell + 1$ in the sense that any α_{t+1} that has positive probability of being drawn at time $t + 1$ satisfies $\alpha_{t+1} \succeq \hat{\alpha}$, $\forall \hat{\alpha} \in \text{supp } \mu_{t-\ell+1}$.

Theorem 4. *If the pace of innovation is fast, then competitive equilibrium is socially optimal, relative to the fixed patent life, ℓ .*

In this case, the positive externality value of investment is 0, and the market outcome coincides with the socially optimal outcome. When ℓ is sufficiently large patents protect technology or knowledge that is worthless, so that $PV_{soc}(\ell)$ is constant as ℓ increases (See figure (3)).

5 Welfare and Patent Length.

The effect of lengthening patent life is to reduce the publicly available technology. What is the impact of such a change on welfare? Because the socially optimal level of investment is higher than that arising in competitive equilibrium, whether lengthening patent length is beneficial or not depends on the impact such changes have on investment. The results to follow identify two cases. When low technology firms are more dependent than high technology firms on the use of technology outside the patent period then, subject to conditions, the impact of lengthening patent life is to force those firms to greater research effort (by depriving them of access to previously unrestricted technology.) And, this has a knock-on effect of increasing the competitive pressure on good firms, forcing them to also raise investment. As a result, overall investment in R&D increases and raises social welfare. In the second case, low technology firms make relatively less use of technology outside the patent period. Then, lengthening the patent period has greater impact on, and inhibits the better technology firms, by reducing the benefits from and incentives to being “good”. In this case the overall impact is to lower social welfare. These results

suggest that patents are beneficial when, *as a result of the need to compete*, they spur R&D and hence innovation. To the extent that disallowing a firm to use the discovery of others ultimately forces that firm to greater investment in R&D the effect of patents is beneficial.

For subsequent discussion, it is useful to write the present value at time t of the payoff flow to a firm, α , optimizing in each period from this point on as: $v(\boldsymbol{\mu}(t), \alpha, \ell)$, where $\boldsymbol{\mu}(t) = (\boldsymbol{\mu}_t, \boldsymbol{\mu}^{t+1})$, so the individual optimization problem may expressed in a Bellman equation as:

$$v(\boldsymbol{\mu}(t), \alpha, \ell) = \max_i \{ \pi(p_t^*, \boldsymbol{\mu}_t, \alpha, \ell) - r(i_t) + \delta \int v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) \} \quad (13)$$

The value function is parametrized by the aggregate distribution sequence, $\boldsymbol{\mu}(t)$, which is determined in equilibrium, but taken as fixed by individual firms. Note that the function v is increasing in α : a firm with higher α can imitate the investment strategy of one with lower α but enjoy lower cost and stochastically better technology draws.

Assuming an interior solution, the first order condition at the solution is:

$$-r'(i) + \delta \lim_{i' \rightarrow i} \left[\frac{1}{i' - i} \right] \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) [P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i', \ell) - P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell)] = 0$$

The first order condition for i is (see the footnote to assumption T3):

$$-r'(i) + \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) = 0. \quad (14)$$

The second order condition for an optimum is then:¹⁹

$$-r''(i) + \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_{ii} P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) < 0. \quad (15)$$

Considering equation (14), $-r'(i) + \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) = 0$, a consequence of assumption T3 is that the optimal value of i increases in α .

In the model developed here, improvements in technology have three effects: demand rises, cost decreases as all firms avail of technology improvements; and each firm becomes more efficient as advances in technology raise the firms ability to make technology improvements. The first two effects unambiguously benefit all firms. The third effect increases competition directly because each firm is more efficient. The following assumption is that the net effect is positive — the value of each firm rises.

- P1. $v(\boldsymbol{\mu}(t), \alpha, \ell)$ is increasing in $\boldsymbol{\mu}^t$: if $\bar{\boldsymbol{\mu}}(t) = (\bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\mu}}_{t+1}, \dots)$ dominates $\hat{\boldsymbol{\mu}}(t) = (\hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{\mu}}_{t+1}, \dots)$ component-wise, then $v(\bar{\boldsymbol{\mu}}(t), \alpha, \ell) \geq v(\hat{\boldsymbol{\mu}}(t), \alpha, \ell)$ for all α .²⁰

¹⁹For the second order condition, assume that $\left[\frac{1}{i' - i} \right] [\Delta_i P(B | \boldsymbol{\mu}_t, \alpha, i', \ell) - \Delta_i P(B | \boldsymbol{\mu}_t, \alpha, i, \ell)]$ converges weakly to a signed measure, $\Delta_{ii} P(\cdot | \boldsymbol{\mu}_t, \alpha, i, \ell)$, as $i' \rightarrow i$.

²⁰In the n -firm linear oligopoly model with demand $P(Q) = a - bQ$ and constant marginal cost, c , profit of firm i is $\pi_i = \left(\frac{1}{n+1} \right)^2 \left(\frac{a-c}{b} \right)^2$, which is increases as c decreases. In the specific context here, improving technology increases demand and supply, but also lowers production cost. If, for example, the net impact is to maintain or raise market price, P1 will be satisfied.

REMARK 5. Recall that the market clearing price is determined by: $p^* = P_d(Q, \boldsymbol{\mu}) = P_s(Q, \boldsymbol{\mu}, \ell)$. If demand depends only on current technology μ_t , and with T1, cost depends only on the distributions μ_t and $\mu_{t-\ell}$, then market clearing price, $p^*(\mu_t, \mu_{t-\ell}, \ell)$, is determined at time t according to $p^* = P_d(Q, \mu_t) = P_s(Q, \mu_t, \mu_{t-\ell}, \ell)$, and profit may be written: $\pi(p^*, \mu_t, \mu_{t-\ell}, \alpha_t, i_i, \ell) = \pi(p^*, \boldsymbol{\mu}_t, \alpha_t, i_i, \ell)$. If (a) $p^*(\mu_t, \mu_{t-\ell}, \ell) \geq p^*(\mu'_t, \mu_{t-\ell}, \ell)$ when $\mu_t \succcurlyeq \mu'_t$, so that better technology pushes up the equilibrium price, and (b) $\pi(p^*, \mu_t, \mu_{t-\ell}, \alpha_t, i_i, \ell) \geq \pi(p^*, \mu_t, \mu'_{t-\ell}, \alpha_t, i_i, \ell)$ when $\mu_{t-\ell} \succcurlyeq \mu'_{t-\ell}$, so that an improvement in cost due to an “across the board” technological improvement raises profit, then P1 is satisfied.

5.1 Patented Knowledge as a Substitute in Cost Reduction and Innovation.

When the (negative) impact of lengthening patent life is greatest on low technology firms, good technology may be considered a substitute for the patented information. Furthermore, if increasing patent length raises the marginal product of investment, then investment also serves as a substitute for patented information. These assumptions are formalized in the following two conditions:

Ia. Increasing patent life impacts lower technology firms profits more. If $\alpha' \succeq \alpha$ then²¹

$$\frac{\partial v(\boldsymbol{\mu}(t), \alpha', \ell)}{\partial \ell} \geq \frac{\partial v(\boldsymbol{\mu}(t), \alpha, \ell)}{\partial \ell}.$$

Ib. Increasing patent life raises the marginal productivity of investment. For g increasing, if $\ell' \geq \ell$,

$$\int_{\tilde{\alpha}} g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}, \alpha, i, \ell') \geq \int_{\tilde{\alpha}} g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}, \alpha, i, \ell).$$

Figure 1 illustrates these assumptions (Since α is not a real number, “ α ” denotes an axis of ordered α 's). An implication of (Ia.) is that the loss of technologies excluded by patents has a greater negative impact on weak or low technology firms. Condition (Ib.) implies that reducing the access of any firm, α , to patented technology raises the marginal productivity of investment by α . Together, these conditions imply that there is greater pressure on weak firms to improve in terms of profitability; and there is greater pay-back to investment after improvement.

When increases in patent life affects good firms less than bad firms and when increased length of patent protection raises the marginal value of investment then increased patent protection raises welfare.

Theorem 5. *Under assumptions T1, T2, T3, P1, Ia and Ib, lengthening patent life improves the aggregate distributions in successive periods and raises the present value of surplus.*

Let PV_S or $PV_S(\ell)$ be the surplus in competitive equilibrium (in the substitutes case.) As ℓ varies, so does the value of surplus, and, when the substitutes condition is satisfied, an increase in ℓ leads to

²¹ $\frac{\partial v(\boldsymbol{\mu}(t), \alpha, \ell)}{\partial \ell}$ measures the marginal impact on the present value of firm α , when ℓ changes but the aggregate distributions on technology over time remain fixed. Recall that $v(\boldsymbol{\mu}(t), \alpha, \ell)$ gives the present value payoff to α optimizing over time with aggregate distributions entering as parameters of the value function.

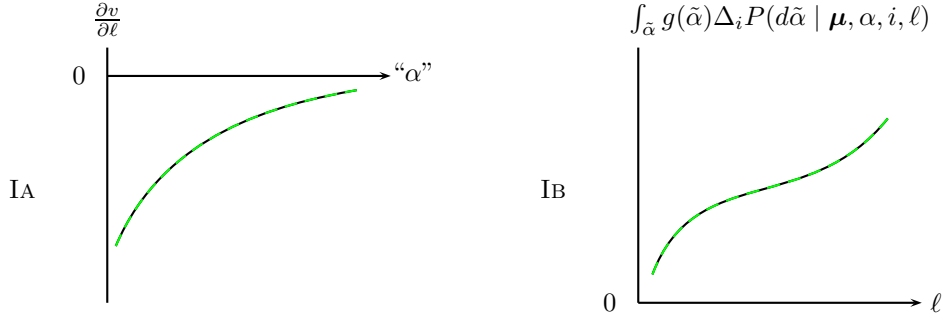


Figure 1: PATENTED KNOWLEDGE SUBSTITUTABLE FOR TECHNOLOGY AND INVESTMENT

an increase in investment and hence raises welfare. The curve PV_S in figure 3 depicts this case. The next section considers the opposite case — where knowledge is complementary to innovation and cost reduction.

5.2 Patented Knowledge as a Complement in Cost Reduction and Innovation.

The effect of lengthening patent life is to reduce the publicly available technology. When the impact of this is greatest on high technology firms, good technology is complemented by the patented information. Furthermore, if increasing patent length reduces the marginal product of investment, then that information is also a complement to investment.

So, in contrast to the previous assumptions (Ia.) and (Ib.), suppose instead that better firms are more dependent on patented information so that such information is a complement to the quality of a firms' technology. Suppose also that increasing patent length removes from use information which raises the marginal product of investment — so that such information is complementary to investment. These conditions are formalized next.

IIa. Increasing patent life impacts higher technology firms profits more. If $\alpha' \succeq \alpha$ then

$$\frac{\partial v(\boldsymbol{\mu}(t), \alpha', \ell)}{\partial \ell} \leq \frac{\partial v(\boldsymbol{\mu}(t), \alpha, \ell)}{\partial \ell}.$$

IIb. Increasing patent life reduces the marginal productivity of investment. For g increasing, $\ell' \geq \ell$,

$$\int_{\tilde{\alpha}} g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}, \alpha, i, \ell') \leq \int_{\tilde{\alpha}} g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}, \alpha, i, \ell).$$

These assumptions are depicted in figure 2.

Under these circumstances, lengthening patent life reduces welfare.

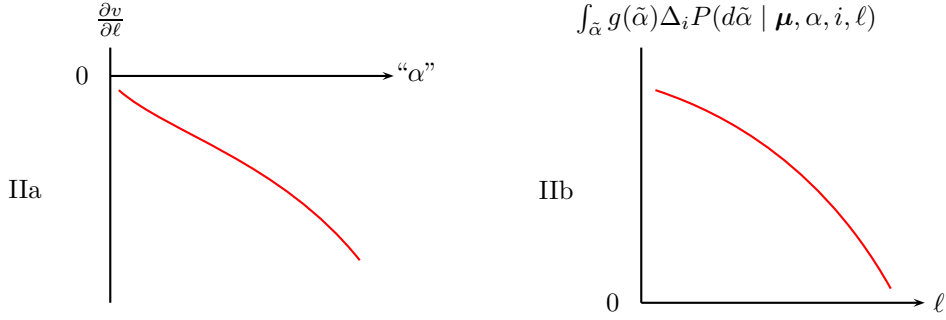


Figure 2: PATENTED KNOWLEDGE COMPLEMENTARY TO TECHNOLOGY AND INVESTMENT

Theorem 6. *Under assumptions T1, T2, T3, P1, IIa and IIb, lengthening patent life worsens the aggregate distributions in successive periods and reduces the present value of surplus.*

As in the earlier discussion write PV_C or $PV_C(\ell)$ to denote the surplus in competitive equilibrium (in the complements case.) In contrast to the substitutes case, here an increase in ℓ leads to a decrease in investment and hence reduces welfare. The curve PV_C in figure 3 depicts this case.

5.3 Welfare comparisons and welfare improvement

Summarizing the previous discussion, when patented information is a substitute for both technology and (the need for) investment then increasing patent life reduces access to that information and forces firms to greater investment, raising social welfare. In the complementary case, the opposite is true: these contrasting cases are shown in figure 3.²²

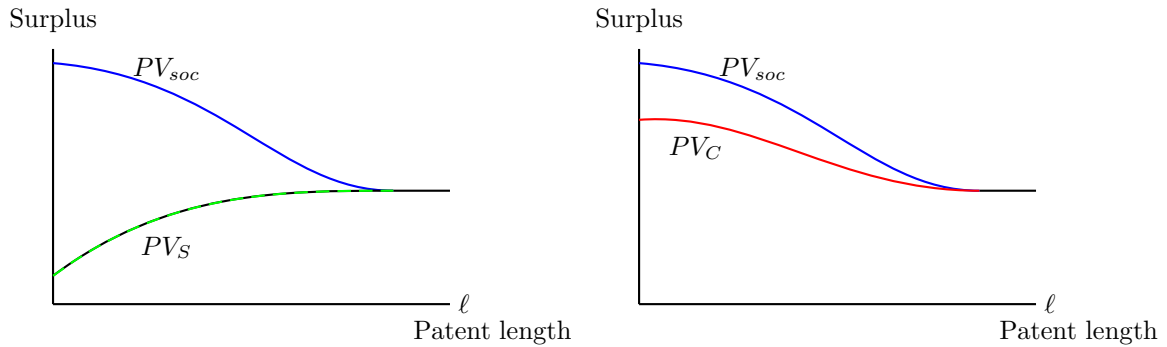


Figure 3: WELFARE AS PATENT LENGTH VARIES

In the substitutes case, increasing patent life raises the value of having good technology (hence encouraging investment indirectly), and raises the direct value of investment in improving one's own

²²When ℓ is very large, patent protection extends beyond obsolescence and changing ℓ has no impact on societal welfare.

technology. Hence, increasing patent length raises welfare. In contrast, with complements, it is advantageous to have good technology at the firm level to benefit from synergies with the available public technology: capitalizing on this synergy encourages investment to improve one's own technology. There, the less available is public technology, the less benefit from private investment. In addition, in the complements case, the direct value of investment is lower since the improvement in own technology is lower when publicly available technology is older. Thus, the benefit of investment is reduced and these effects together imply that lengthening patent life reduces welfare.

When these effects conflict the consequence of increasing patent life is ambiguous. This occurs, for example, when increasing patent life has greater (negative) impact on better technology firms but at the same time increases the marginal productivity of investment (in generating innovation). In this case, there is less incentive to improve, but it is easier to do so.

One special case of interest is that where the impact of changing patent length on the marginal productivity is small or 0:

$$\frac{1}{\ell' - \ell} \left[\int_{\tilde{\alpha}} g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}, \alpha, i, \ell') - \int_{\tilde{\alpha}} g(\tilde{\alpha}) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}, \alpha, i, \ell) \right] \approx 0.$$

In this case, if increasing patent length has a more adverse effect on weaker (low technology) firms, then it is welfare improving; and if the effect is greater on stronger firms, then it is welfare reducing. In particular, patent policy plays a beneficial role when it *forces* less technologically advanced firms to invest more (by depriving them of the use of others ideas). So, patents are beneficial not because they *encourage* reward seeking behavior, but because to survive a firm is compelled to invest and innovate: competitive pressure rather than the prize of monopoly spurs research. To the extent that one may extrapolate from these observations a scheme that gives (payoff) advantage to weak innovators is bad from a welfare perspective. On the other hand, patents are detrimental to welfare when the direct negative effect from lengthening patent life is greatest for the better technology firms, since it reduces the incentive to invest.

5.3.1 Welfare Improvement

Standard tax-subsidy schemes provide welfare improving incentives (with the implicit presumption that policy makers have full information on the costs and benefits of innovation). Given a prevailing patent regime, provided innovation is not fast relative to patent life, an investment subsidy and lump sum tax can improve welfare. The follow result makes this assertion at a patent life of 0.

Theorem 7. *Suppose that $\ell = 0$. Then there is an individual specific pricing scheme that raises welfare and is self financing.*

However, in practice such schemes may not be practical and are not common as a welfare improving device. The central problem resulting from the delay in having discovery widely used is that the gain to a firm from holding exclusive rights to discovery, is less than the potential gain to society from having immediate access to the discovery. The compensation required to give the developing firm

the necessary incentives is less than the overall value to society. In principle all firms together would be willing, as a group, to pay sufficient compensation to reward the discoverer's effort in return for access to the discovery. How can this be achieved? The most commonly discussed method of making patented discovery available is through licensing: the discoverer makes available use of the innovation for a fee, and the revenue obtained in this way outweighs cost from loss of control over the use of the innovation. Licensing is incorporated in [13] so that when licensing is necessary because of infringement the incremental profit is divided so that all involved parties have non-negative share. There, innovation improves quality, q , with a discovery producing a jump, Δ , in quality. In [10] the holdup potential of licensing is examined.

In the framework developed here licensing is complex because the value of innovation from period to period is correlated across firms: it is the overall improvement that matters, rather than any individual discovery. And because technology is multidimensional, different firms may add value along different dimensions so that the valuation of the individual contribution to overall discovery from period to period is difficult or impossible to ascertain. In this case licensing may be difficult to implement since it is a "batch" of discoveries rather than any specific one that matters and there is no natural bargaining mechanism for firms to resolve such issues, especially when the number of firms is large. This observation may well be reflected in practice with the increasing litigation over ownership of ideas.

6 Conclusion

The traditional argument for patents is that they encourage innovation by giving the innovator monopoly power for a period of time: monopoly rights create the incentive to invest so that innovation in the aggregate is greater than it would be in the absence of patents. In the environment here, individual innovations are important to the firm but alone not significant in the overall pool of discovery from period to period. The synergy from the pooling of innovation is what creates the externality value in discovery. But the patent blocks the innovator from benefiting from that pool of discovery. While the patent system cannot achieve an efficient outcome for the market — because some of the positive externalities cannot be internalized, this blocking effect can have either a positive or a negative impact on overall innovation and welfare. Whether patents improve or worsen overall welfare depends on the exact way in which innovation interacts with investment and firm quality (as discussed in (5.1) and (5.2)).

This discussion focuses entirely on the incentive effects that arise from obtaining monopoly rights on innovation when there are large numbers of innovators. In particular, the prospect of licensing is not considered. One might argue that somehow the positive externalities might be internalized by the firms in the market. However, as mentioned above, there are good reasons to expect that this might not happen. Large numbers of innovators make reaching consensus on sharing of innovation difficult, but beyond this, the overlap in discovery within period and the dependent evolution of discovery over time make it difficult if not impossible to price individual innovations and hence allocate value to ideas. So, while possibility exists that a market might develop to price each innovator's discoveries, it is questionable

as to whether or not this would occur.

Appendix I: Evolution of technology distributions

With a view to developing a recursion, recall from equations (2) and (3):

$$\mu_{t+1}(\cdot) = \varphi_{t,1}(\cdot \mid \boldsymbol{\mu}_t, \tau_t, \ell) = \varphi_{t,1}(\cdot \mid \boldsymbol{\mu}_t, \boldsymbol{\tau}^t, \ell), \quad (16)$$

where for notational convenience, $\boldsymbol{\tau}^t = (\tau_t, \tau_{t+1}, \dots)$, is the sequence of distributional strategies going forward in time, and where $\varphi_{t,1}(\cdot \mid \boldsymbol{\mu}_t, \boldsymbol{\tau}^t, \ell)$ has $\boldsymbol{\tau}^t$ as an argument to allow for dependence of the j period ahead aggregate distribution on values of $\boldsymbol{\tau}^t$ beyond t . Similarly,

$$\psi_{t,1}(\cdot \mid \boldsymbol{\mu}_t, \boldsymbol{\tau}^t, \alpha_t, i_t, \ell) = P(\cdot \mid \boldsymbol{\mu}_t, \alpha_t, i_t, \ell). \quad (17)$$

So, $\psi_{t,1}$ gives the distribution over technology that firm α_t , investing i_t , will draw from next period. Aggregating across individual firms yields the aggregate distribution on technologies.

$$\mu_{t+1}(\cdot) = \varphi_{t,1}(\cdot \mid \boldsymbol{\mu}_t, \tau_t, \ell) = \int_{\alpha_t, i_t} \psi_{t,1}(\cdot \mid \boldsymbol{\mu}_t, \boldsymbol{\tau}^t, \alpha_t, i_t, \ell) \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \quad (18)$$

From equation (2 or 16), observe that μ_{t+1} is determined by $\boldsymbol{\mu}_t$ and τ_t . Going forward, μ_{t+1} is determined by $\boldsymbol{\mu}_t$, τ_t and τ_{t+1} . And so on. Let $\boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t) = \boldsymbol{\mu}_{t+j} = (\mu_t, \mu_{t+1}, \dots, \mu_{t+j-1}, \mu_{t+j})$, with $\mu_t = \text{marg}_{\mathcal{A}} \tau_t$, μ_{t+1} determined according to the iteration in equation (16), and $\{\mu_{t+j}\}_{j>1}$ defined inductively. Going forward one period, define

$$\begin{aligned} \psi_{t,2}(\cdot \mid \boldsymbol{\mu}_{t+1}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \alpha_t, i_t, \ell) \\ = \int_{\alpha_{t+1}, i_{t+1}} P(\cdot \mid \boldsymbol{\mu}_{t+1}, \alpha_{t+1}, i_{t+1}, \ell) \tau_{t+1}(di_{t+1} \mid \alpha_{t+1}) \psi_{t,1}(\cdot \mid \boldsymbol{\mu}_t, \boldsymbol{\tau}^t, \alpha_t, i_t, \ell) \end{aligned} \quad (19)$$

The distribution $\psi_{t,2}$ gives the the two-period ahead distribution over characteristics for firm α_t investing i_t currently and following τ_{t+1} next period. Averaging:

$$\mu_{t+2}(\cdot) = \varphi_{t,2}(\cdot \mid \boldsymbol{\mu}_{t+1}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \ell) = \int_{\alpha_t, i_t} \psi_{t,2}(\cdot \mid \boldsymbol{\mu}_{t+1}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \alpha_t, i_t, \ell) \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \quad (20)$$

Write $\mu_{t+j+1}(\cdot) = \varphi_{t,j+1}(\cdot \mid \boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \ell)$ to denote the aggregate distribution at time $t+j+1$ given $\boldsymbol{\mu}_t$ and $\boldsymbol{\tau}^t$; and $\psi_{t,j+1}(\cdot \mid \boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \alpha_t, i_t, \ell)$ for the distribution over technology of firm α , $j+1$ periods after t conditional on $\boldsymbol{\mu}_t$ and $\boldsymbol{\tau}^t$. This gives the distribution at time $t+j+1$ over technologies for firm α investing i at time t and following the investment strategy $\tau_t(\cdot \mid \cdot)$ thereafter. The $j+1$ period ahead individual distribution is:

$$\begin{aligned} \psi_{t,j+1}(\cdot \mid \boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \alpha_t, i_t, \ell) = \\ \int_{\alpha_{t+j}, i_{t+j}} P(\cdot \mid \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, i_{t+j}, \ell) \tau_{t+j}(di_{t+j} \mid \alpha_{t+j}) \psi_{t,j}(d\alpha_{t+j} \mid \boldsymbol{\mu}_t, \boldsymbol{\tau}^t, \alpha_t, i_t, \ell), \end{aligned} \quad (21)$$

and the aggregate and individual distributions are related by the formula:

$$\mu_{t+j+1}(\cdot) = \varphi_{t,j+1}(\cdot \mid \boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \ell) = \int_{\alpha_t, i_t} \psi_{t,j+1}(\cdot \mid \boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \alpha_t, i_t, \ell) \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \quad (22)$$

Appendix II: Proofs

Proof of Theorem 1: Recall equation 8 in the text:

$$\begin{aligned}
V(\boldsymbol{\mu}, \ell) &= \max_{\boldsymbol{\tau} \in C^*(\boldsymbol{\mu})} \sum_{t=1}^{\infty} \delta^{t-1} [S(\boldsymbol{\mu}_t, \ell) - \int r(i) d\tau_t] \\
&= \max_{\boldsymbol{\tau} \in C^*(\boldsymbol{\mu})} \sum_{t=1}^{\infty} \delta^{t-1} [S(\boldsymbol{\mu}_t(\boldsymbol{\tau}^t), \ell) - \int r(i) d\tau_t] \\
&= \max_{\boldsymbol{\tau} \in C^*(\boldsymbol{\mu})} \hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)
\end{aligned} \tag{23}$$

where $\boldsymbol{\mu}$ is defined inductively from $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ and $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots)$, according to equations (2) and (18), and where $\hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)$ is the present value of the surplus flow from the strategy $\boldsymbol{\tau}$ with initial distribution $\boldsymbol{\mu} = \boldsymbol{\mu}_1$.

Varying τ_t in the direction $\Delta\tau_t = (\tau'_t - \tau_t)$: $\tilde{\tau}_t = \tau_t + \epsilon\Delta\tau_t = (1 - \epsilon)\tau_t + \epsilon\tau'_t$. It is necessary to show that:

$$\begin{aligned}
\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\hat{V}(\boldsymbol{\mu}, \tilde{\boldsymbol{\tau}}, \ell) - \hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)] &= \tag{24} \\
&\delta^{t-1} \cdot \sum_{j=1}^{\infty} \delta^j \int_{z_t} \left\{ \partial_C \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, l \mid \Delta\tau_t) + \partial_T \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, l \mid \Delta\tau_t) \right\} \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \\
&+ \delta^{t-1} \left\{ \sum_{j=1}^{\infty} \delta^j \left[\int_{z_t} \bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, l) - \int_{z_{t+j}} r(i_{t+j}) \tau_{t+j}(di_{t+j} \mid \alpha_{t+j}) \psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t, l) \right. \right. \\
&\quad \left. \left. - \int_{z_t} r(i) \right\} \Delta\tau_t(di \mid \alpha_t) \mu_t(d\alpha_t)
\end{aligned}$$

At periods $t + 1$, $t + 2$, and so on, the perturbation impacts the aggregate distribution directly; and from period $t + l$ the transition kernel and cost function are affected as the period t technology becomes publicly available. The surplus variation at time $t + j$ due to the variation in τ_t is $S(\tilde{\boldsymbol{\mu}}_{t+j}, \ell) - S(\boldsymbol{\mu}_{t+j}, \ell)$. The following calculations consider the value of:

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [S(\tilde{\boldsymbol{\mu}}_{t+j}, \ell) - S(\boldsymbol{\mu}_{t+j}, \ell)].$$

For the computations to follow, simplify notation by writing $\psi_{t,j+1}$ or $\psi_{t,j+1}(d\alpha_{t+j+1} \mid \alpha_t, i_t)$ in place of $\psi_{t,j+1}(d\alpha_{t+j+1} \mid \boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \alpha_t, i_t, \ell)$ from equation (21). Similarly, write $\varphi_{t,j+1}$ in place of $\varphi_{t,j+1}(\cdot \mid \boldsymbol{\mu}_{t+j}(\boldsymbol{\tau}^t), \boldsymbol{\tau}^t, \ell)$ in equation (22). At period $t + l$, $\tilde{\boldsymbol{\mu}}_{t+l} = (\tilde{\mu}_{t+l}, \dots, \tilde{\mu}_{t+1} \mid \boldsymbol{\mu}_t)$. The effect of changing τ_t is direct from the next period, $t + 1$, through $\tilde{\mu}_{t+1}$ and subsequent aggregate distributions (according to equations (18) and (22)).

REMARK 6. Note that at time t ,

$$\mu_{t+1} = \int_{\alpha_t, i_t} P(\cdot \mid (\mu_t, \dots, \mu_{t-l+1} \mid \mu_{t-l}, \dots), \alpha_t, i_t) \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \tag{25}$$

depends only on the aggregate distribution at time $t - l$, the most recent vintage of technology that is freely usable, and which has a role in determining the aggregate distribution at time $t + 1$. In a similar manner, the action variation $\tau_t \rightarrow \tilde{\tau}_t$ changes μ_{t+1} to $\tilde{\mu}_{t+1}$, and this enters as a parameter to affect the distribution at period $t + l + 1$ through the transition kernel.

Considering the variation on the supply side at time $t + j$, since the demand curve does not shift, the change in area between the curves is approximated by:

$$\begin{aligned} \Delta S_{t+j} &= S(\boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}), \ell) - S(\boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1}), \ell) \\ &\approx \int_0^{p^*} Q_s(p_{t+j}, (\boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}), \ell) dp_{t+j} - \int_0^{p^*} Q_s(p, (\boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1}), \ell) dp_{t+j}, \end{aligned} \quad (26)$$

where p_{t+j}^* is the market clearing price at aggregate distribution $\varphi_{t,j}$. Write $\Delta\tau_t = \tau_t' - \tau_t$ and $\tilde{\tau}_t = \tau_t + \epsilon\Delta\tau_t$. Recalling equations (21) and (22), for $j \geq 1$, let

$$\tilde{\psi}_{t,j+1} = \psi^{j+1}(\cdot | \boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}), (\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}), \alpha_t, i_t, \ell); \quad \tilde{\varphi}_{t,j+1} = \tilde{\varphi}^{j+1}(\cdot | \boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}), \tilde{\tau}_t, \boldsymbol{\tau}^{t+1}, \ell) \quad (27)$$

and put

$$\Delta\varphi_{t,j} = \tilde{\varphi}_{t,j} - \varphi_{t,j}, \quad \Delta\psi_{t,j} = \tilde{\psi}_{t,j} - \psi_{t,j}.$$

Finally, write $\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) = \psi_{t,j}(d\alpha_{t+j} | \tau_t, \alpha_t, i_t)$, and $\tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) = \psi_{t,j}(d\alpha_{t+j} | \tilde{\tau}_t, \alpha_t, i_t)$. Then,

$$\begin{aligned} \Delta S_{t+j} &\approx \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \alpha_{t+j}, \boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}), \ell) \tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tilde{\tau}_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp - \\ &\quad \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \alpha_{t+j}, \boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1}), \ell) \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \end{aligned}$$

For simplicity, write $q(p_{t+j}, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) = q(p, \boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}), \alpha_{t+j}, \ell)$ to denote the time $t + j$ supply of firm α_{t+j} at price p , given the time t aggregate distribution τ_t . Similarly, write $\tilde{\boldsymbol{\mu}}_{t+j} = \boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1})$, and $\tilde{\mu}_{t+j} = \mu_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1})$. And, taking $\boldsymbol{\tau}^{t+1}$ as fixed throughout the discussion so that replacing $\tilde{\tau}_t = \tau_t + \epsilon\Delta\tau_t$, $\tilde{\tau}_t(di_t | \alpha_t) = \tau_t(di_t | \alpha_t) + \epsilon\Delta\tau_t(di_t | \alpha_t)$:

$$\begin{aligned} \Delta S_{t+j} &\approx \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \\ &\quad + \epsilon \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \Delta\tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \\ &\quad - \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \end{aligned}$$

Rearranging with $\tilde{\psi}_{t,j} \stackrel{\text{def}}{=} \psi_{t,j} + \Delta\psi_{t,j}$,

$$\begin{aligned} \Delta S_{t+j} &\approx \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) [\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t)] \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \\ &\quad + \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \Delta\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \\ &\quad - \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \\ &\quad + \epsilon \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \Delta\tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \end{aligned}$$

Rearranging again,

$$\begin{aligned} \Delta S_{t+j} &\approx \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} [q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) - q(p, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell)] \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \\ &\quad + \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \Delta\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \\ &\quad + \epsilon \int_0^{p_{t+j}^*} \int_{z_t} \int_{\alpha_{t+j}} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \Delta\tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) dp \end{aligned}$$

Changing the order of integration:

$$\begin{aligned} \Delta S_{t+j} &\approx \int_{z_t} \int_{\alpha_{t+j}} \int_0^{p_{t+j}^*} [q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) - q(p, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell)] dp \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) + \\ &\quad \int_{z_t} \int_{\alpha_{t+j}} \int_0^{p_{t+j}^*} q(p, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) dp \Delta\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \\ &\quad + \epsilon \int_{z_t} \int_{\alpha_{t+j}} \int_0^{p_{t+j}^*} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) dp \tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \Delta\tau_t(di_t | \alpha_t) \mu_t(d\alpha_t). \end{aligned}$$

Using the fact that

$$\int_0^{p_{t+j}^*} q(p, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) dp = \int_0^{q_{t+j}^*} [p_{t+j}^* - c_q(q, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell)] dq = \pi(p_{t+j}^*, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell),$$

where q_{t+j}^* sets marginal cost equal to p_{t+j}^* gives:

$$\begin{aligned} \Delta S_{t+j} &\approx \int_{z_t} \int_{\alpha_{t+j}} [\pi(p_{t+j}^*, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) - \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell)] \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \\ &\quad + \int_{z_t} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \Delta\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \\ &\quad + \epsilon \int_{z_t} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) \tilde{\psi}_{t,j}(d\alpha_{t+j} | \alpha_t, i_t) \Delta\tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \end{aligned} \quad (28)$$

Dividing by ϵ and letting $\epsilon \rightarrow 0$ assuming that $\Delta\psi_{t,j}$ is of order ϵ , gives:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \Delta S_{t+j} &= \int_{z_t} \int_{\alpha_{t+j}} [D\pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, l \mid \Delta\tau_t) \psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t) \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \\ &\quad + \int_{z_t} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \delta\psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t) \tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \\ &\quad + \int_{z_t} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t) \Delta\tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t)] \end{aligned} \quad (29)$$

where, recalling $\boldsymbol{\mu}_{t+j} = \boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1})$ and $\tilde{\boldsymbol{\mu}}_{t+j} = \boldsymbol{\mu}_{t+j}(\tilde{\tau}_t, \boldsymbol{\tau}^{t+1}) = \boldsymbol{\mu}_{t+j}(\tau_t + \epsilon\Delta\tau_t, \boldsymbol{\tau}^{t+1})$,

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\pi(p_{t+j}^*, \tilde{\boldsymbol{\mu}}_{t+j}, \alpha_{t+j}, \ell) - \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell)] \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\pi(p_{t+j}^*, \tau_t + \epsilon\Delta\tau_t, \alpha_{t+j}, \ell) - \pi(p_{t+j}^*, \alpha_{t+j}, \tau_t, \ell)] \\ &= D\pi(p_{t+j}^*, \tau_t, \alpha_{t+j}, l \mid \Delta\tau_t), \end{aligned}$$

and where

$$\int_{\alpha_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \delta\psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t) \stackrel{\text{def}}{=} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \Delta\psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t).$$

Consider the terms in (28) in turn. Define

$$\partial_C \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell \mid \Delta\tau_t) \stackrel{\text{def}}{=} \int_{\alpha_{t+j}} D\pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell \mid \Delta\tau_t) \psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t, \ell). \quad (30)$$

This gives the expected variation in profit to α_t resulting from cost reduction at time $t+j$ arising from the innovation generated by the *aggregate distribution shift* $\Delta\tau_t$ as the discovery becomes publicly usable.

Similarly, let $\partial_T \pi_{t+j}$ denote the impact on profit of an expected improvement in technology draw at time $t+j$ resulting from the perturbation of the transition kernel by the *shift in the aggregate distribution*:

$$\partial_T \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell \mid \Delta\tau_t) \stackrel{\text{def}}{=} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \delta\psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t, \ell) \quad (31)$$

The remaining term gives the *direct effect* on period $t+j$ profit of the investment change in period t , denoted $\partial_D \pi_{t+j}$ and defined:

$$\bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell) \stackrel{\text{def}}{=} \int_{\alpha_{t+j}} \pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) \psi_{t,j}(d\alpha_{t+j} \mid \alpha_t, i_t, \ell) \quad (32)$$

With this notation,

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \Delta S_{t+j} &= \int_{z_t} \left\{ \partial_C \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell \mid \Delta\tau_t) + \partial_T \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell \mid \Delta\tau_t) \right\} \Delta\tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \\ &\quad + \int_{z_t} \bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell) \Delta\tau_t(di_t \mid \alpha_t) \mu_t(d\alpha_t) \end{aligned}$$

The remaining effect to be considered is the direct effect on future investment. At period $t + j$, this is given by: $\int_{z_{t+j}} r(i_{t+j})\tau_{t+j}(di_{t+j} | \alpha_{t+j})\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, l)$. Taking expectations of terms individually and gathering terms:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\hat{V}(\boldsymbol{\mu}, \tilde{\boldsymbol{\tau}}, \ell) - \hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)] = & \quad (33) \\ & \delta^{t-1} \cdot \sum_{j=1}^{\infty} \delta^j \int_{z_t} \left\{ \partial_C \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t) + \partial_T \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t) \right\} \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \\ & + \delta^{t-1} \left\{ \sum_{j=1}^{\infty} \delta^j \int_{z_t} \bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell) - \int_{z_t} \int_{z_{t+j}} r(i_{t+j})\tau_{t+j}(di_{t+j} | \alpha_{t+j})\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, l) \right. \\ & \quad \left. - \int_{z_t} r(i_t) \right\} \Delta\tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \end{aligned}$$

Where $\boldsymbol{\tau} = (\tau_1(\cdot | \alpha), \dots, \tau_t(\cdot | \alpha), \dots)$ and $\tilde{\boldsymbol{\tau}}$ is defined from $\boldsymbol{\tau}$ by replacing $\tau_t(\cdot | \alpha)$ with $\tilde{\tau}_t(\cdot | \alpha)$. Rearranging the last terms:

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\hat{V}(\boldsymbol{\mu}, \tilde{\boldsymbol{\tau}}, \ell) - \hat{V}(\boldsymbol{\mu}, \boldsymbol{\tau}, \ell)] = & \quad (34) \\ & \delta^{t-1} \cdot \sum_{j=1}^{\infty} \delta^j \int_{z_t} \left\{ \partial_C \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t) + \partial_T \pi_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell | \Delta\tau_t) \right\} \tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \\ & + \delta^{t-1} \sum_{j=1}^{\infty} \delta^j \int_{z_t} \left\{ \bar{\pi}_{t+j}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, i_t, \ell) - \int_{z_{t+j}} r(i_{t+j})\tau_{t+j}(di_{t+j} | \alpha_{t+j})\psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, l) \right. \\ & \quad \left. - r(i_t) \right\} \Delta\tau_t(di_t | \alpha_t) \mu_t(d\alpha_t) \end{aligned}$$

This gives the variation appearing in the statement of the theorem. Optimality requires that all feasible variations produce non-positive surplus variation. ■

Proof of theorem 2: Profit is (weakly) declining in ℓ because cost is weakly increasing in ℓ , and for $\ell' \geq \ell$ the transition kernel satisfies $P(\cdot | \boldsymbol{\mu}, \alpha, i, \ell') \succcurlyeq P(\cdot | \boldsymbol{\mu}, \alpha, i, \ell)$ in first order stochastic dominance. From the former, period surplus declines (other things equal), and from the latter, other things equal, the aggregate technology distribution is worse each period, moving the supply inward. Considering the change in surplus from a variation in ℓ (instead of τ_t), keeping all else constant:

$$\begin{aligned} \Delta S_{t+j} &= S(\boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1}), \tilde{\ell}) - S(\boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1}), \ell) \\ &\approx \int_0^{p^*} Q_s(p_{t+j}, (\boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1}), \tilde{\ell}) dp_{t+j} - \int_0^{p^*} Q_s(p, (\boldsymbol{\mu}_{t+j}(\tau_t, \boldsymbol{\tau}^{t+1}), \ell) dp_{t+j}. \end{aligned} \quad (35)$$

If $\tilde{\ell} < \ell$, then $\Delta S_{t+j} > 0$ — because with an unchanged investment strategy individual output will be higher and aggregate supply pushed out, increasing surplus. Thus, at $\tilde{\ell}$ surplus must be higher when the present value of surplus is optimized at $\tilde{\ell}$. Hence, lowering ℓ lowers surplus and so the optimal value

of ℓ is 0. ■

Proof of theorem 3: From equation 12 optimal period t investment for firm α solves:

$$\max_{i_t} \left\{ \sum_{j=1}^{\infty} \delta^j [\bar{\pi}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, \ell) - \int_{z_{t+j}} r(i_{t+j}) \tau_{t+j}(di_{t+j} | \alpha_{t+j}) \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, \ell)] - r(i_t) \right\} \quad (36)$$

giving the first order condition:

$$-r'(i_t) + \sum_{j=1}^{\infty} \delta^j \int_{\alpha_{t+j}} [\pi(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_{t+j}, \ell) - r(i_{t+j}) \tau_{t+j}(di_{t+j} | \alpha_{t+j})] \Delta_i \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, \ell) = 0. \quad (37)$$

If $i_t(\alpha_t)$ solves this expression, define $\hat{\tau}_t$ as the measure with marginal μ_t and support on the set $\{(i_t, \alpha_t) | i_t = i_t(\alpha_t)\}$. Considering the optimization problem:

$$\max_{\tau_t} \int_{z_t} \left[\sum_{j=1}^{\infty} \delta^j \left\{ [\bar{\pi}(p_{t+j}^*, \boldsymbol{\mu}_{t+j}, \alpha_t, \ell) - \int_{z_{t+j}} r(i_{t+j}) \tau_{t+j}(di_{t+j} | \alpha_{t+j}) \psi_{t,j}(d\alpha_{t+j} | \alpha_t, i_t, \ell)] - r(i_t) \right\} \right] \tau(di_t | \alpha_t) \mu(d\alpha_t) \quad (38)$$

the measure $\hat{\tau}_t$ is a solution. Conversely, if $\bar{\tau}_t$ maximizes (38), then for μ_t -almost all α_t , almost all i_t in the support of $\bar{\tau}_t(\cdot | \alpha_t)$ maximize (36).

In the market equilibrium, firms calculate the benefit of innovation ignoring the long-term effect on aggregate innovation and cost. The social benefit of investment exceeds the private benefit by the first two terms in equation (10). Comparing expressions (10) and (36), when $\partial_C \pi_{t+j} = \partial_T \pi_{t+j} = 0$ in expression (10), both have the same solution. ■

Proof of theorem 4: In view of the previous theorem, the terms $\partial_C \pi_{t+j}$ and $\partial_T \pi_{t+j}$ are 0 in the social planner optimization in equation (10). ■

Proof of theorem 5: Considering equation (14):

$$-r'(i) + \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) \equiv 0$$

the marginal impact on firm α_t (ignoring the impact on equilibrium of a variation in ℓ) is obtained by

differentiating with respect to ℓ .

$$\begin{aligned}
-r''(i) \frac{di_t}{d\ell} + \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_{ii} P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) \frac{di_t}{d\ell} \\
+ \delta \int_{\tilde{\alpha}} v_\ell(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) \\
+ \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_{ii} P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) \equiv 0
\end{aligned} \tag{39}$$

Therefore:

$$\frac{di_t}{d\ell} = \frac{\delta \int_{\tilde{\alpha}} v_\ell(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) + \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_{ii} P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell)}{r''(i_t) - \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t+1), \tilde{\alpha}, \ell) \Delta_{ii} P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell)} \tag{40}$$

From the second order condition, the denominator is positive, so the sign of $\frac{di_t}{d\ell}$ is the same as that of the numerator. Assumption Ib implies that $\int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell) \Delta_{ii} P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) > 0$ and in view of T2, Ia implies that $\int_{\tilde{\alpha}} v_\ell(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell) \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) > 0$. Consequently using (Ia,b), $\frac{di_t}{d\ell} > 0$.

These calculations ignore the impact of changes in investment behavior on the aggregate distribution. Recall the first order condition:

$$r'(i) = \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell) \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell).$$

After raising ℓ , the variation in i_t , holding fixed $\boldsymbol{\mu}(t)$ is upward. With $\ell' > \ell$, $i' > i$.

$$r'(i') = \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell') \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i', \ell').$$

However, higher investment will impact $\boldsymbol{\mu}_t$, raising the quality of the aggregate distribution next period. From T3, better technology firms invest more, so that the distribution is better in subsequent periods. This leads to a (weak) decrease in each firms' cost and an improvement in the technology draw next period. Both of these effects work to raise v so that the resulting value function satisfies:

$$v(\boldsymbol{\mu}'(t), \tilde{\alpha}, \ell') > v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell').$$

In this case,

$$r'(i') = \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell') \Delta_i P(d\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i', \ell') < \delta \int_{\tilde{\alpha}} v(\bar{\boldsymbol{\mu}}'(t), \tilde{\alpha}, \ell') \Delta_i P(d\tilde{\alpha} | \bar{\boldsymbol{\mu}}'^t, \tilde{\alpha}, i', \ell').$$

To restore equality, i must rise (further) with consequent (further) impact on the aggregate distribution. Assuming $r'(x)$ is sufficiently large for large values of x , the iterative process will eventually converge to equilibrium. Consequently the impact of increasing ℓ is to raise the aggregate distribution quality in subsequent periods and hence the present value of surplus (welfare). ■

Proof of theorem 6: Assumption IIb implies that $\int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell) \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) < 0$ and in view of T2, Ia implies that $\int_{\tilde{\alpha}} v_\ell(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell) \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell) < 0$. Consequently using (IIa,b), $\frac{di_t}{d\ell} < 0$.

As before, these calculations ignore the impact of changes in investment behavior on the aggregate distribution. So, reconsider the first order condition:

$$r'(i) = \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell) \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i, \ell).$$

After raising ℓ , the variation in i_t , holding fixed $\bar{\boldsymbol{\mu}}^t$ is downward. With $\ell' > \ell$, $i' < i$.

$$r'(i') = \delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell') \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i', \ell').$$

Again, this expression ignores the fact that lower investment will impact $\boldsymbol{\mu}_t$, reducing the quality of the aggregate distribution in future periods. This leads to a (weak) increase in each firms' cost and a disimprovement in the technology draw next period. Both of these effects work to reduce v so that the resulting value function satisfies

$$v(\boldsymbol{\mu}'(t), \tilde{\alpha}, \ell') < v(\boldsymbol{\mu}(t), \tilde{\alpha}, \ell').$$

In this case,

$$\delta \int_{\tilde{\alpha}} v(\boldsymbol{\mu}'(t), \tilde{\alpha}, \ell') \Delta_i P(\bar{\boldsymbol{\mu}}^t, \tilde{\alpha}, i', \ell') < \delta \int_{\tilde{\alpha}} v(\bar{\boldsymbol{\mu}}^t, \tilde{\alpha}, \ell') \Delta_i P(\tilde{\alpha} | \boldsymbol{\mu}_t, \alpha, i', \ell') = r'(i').$$

To restore equality, i must fall (further) with consequent (further) impact on the aggregate distribution. If $r''(0)$ is sufficiently small, the iterative process will eventually converge to equilibrium. Consequently the impact of increasing ℓ is to worsen the aggregate distribution quality in subsequent periods and hence reduce the present value of surplus (welfare). ■

Proof of theorem 7: Consider a deterministic investment variation with $\Delta\tau_t = \tau'_t - \tau_t$ and with τ'_t and τ_t having degenerate conditional distributions at each α_t : $\tau'_t(di | \alpha_t)$ has support $i'(\alpha_t)$ and τ_t has support $i(\alpha_t)$ — so that $\Delta\tau_t(di | \alpha_t) = i'(\alpha_t) - i(\alpha_t)$. Also, take $\tau'_t(di | \alpha_t)$ to have 0 mass outside a set A^* — only investment of firms or technologies in the set A^* are perturbed. For firm $\alpha_t \in A^*$, define an investment subsidy

$$\rho_t(\alpha_t) = \begin{cases} \sum_{j=1}^{\infty} \delta^j \left\{ \partial_{C+T} \pi_{t+j}(p_{t+j}^*, \tau_t, \alpha_t, i(\alpha_t), \ell | \Delta\tau_t) \right\} & \alpha_t \in A^* \\ 0, & \alpha_t \notin A^* \end{cases}$$

The subsidy pays $\rho(\alpha_t)$ per unit if investment is between $i(\alpha_t)$ and $i'(\alpha_t)$ and 0 otherwise. In addition, firm α_t pays a fixed cost of $f(\alpha_t) = \rho(\alpha_t)[i'(\alpha_t) - i(\alpha_t)]$. With this scheme, firm $\alpha_t \in A^*$ optimizes with investment equal to $i'(\alpha_t)$. Given the fixed cost, the welfare of firms in A^* is unchanged. However, with

$(A^*)^c$ the complement of A^*), since

$$\int_{(A^*)^c} \sum_{j=1}^{\infty} \delta^j \int_{i_t} \left\{ \partial_{C+T} \pi_{t+j}(p_{t+j}^*, \tau_t, \alpha_t, i_t, \ell \mid \Delta \tau_t) \right\} \tau_t(di_t \mid \alpha_t) > 0, \quad (41)$$

total welfare is raised.

■

Appendix III: Dominance

For the following review, take as given: (a1) Λ , a completely regular topological space (for example, Λ a metric space), (a2) \mathcal{B}_Λ the Borel field on Λ , (b) \succeq , an order on Λ (reflexive, transitive and antisymmetric), (c) $C_b(\Lambda)$, the set of continuous bounded real-valued functions on Λ , (d) $\mathcal{M}_+(\Lambda)$, the set of non-negative measures on Λ , and (e) $\mathcal{P}(\Lambda)$ the set of probability measures on Λ . (A topological space Λ is completely regular if and only if when A is closed in Λ and $\alpha \notin A$, there is a continuous function, f , $f : \Lambda \rightarrow [0, 1]$ such that $f(\alpha) = 0$ and $f(A) = 1$.)

Definition 2. A real valued function $f : \Lambda \rightarrow \mathfrak{R}$ is called increasing if $\alpha' \succeq \alpha$ implies that $f(\alpha') \geq f(\alpha)$ (and decreasing if $\alpha' \succeq \alpha$ implies that $f(\alpha') \leq f(\alpha)$). Write $\mathcal{I}_m(\Lambda)$ for the set of increasing measurable functions on Λ .

A set $B \subseteq \Lambda$ is called increasing if $x, y \in \Lambda$, $x \in B$ and $y \succeq x$ imply that $y \in B$.

Definition 3. Given $\mu, \nu \in \mathcal{P}(\Lambda)$, define a pre-ordering (reflexive and transitive relation) on $\mathcal{P}(\Lambda)$:

$$\mu \succeq \nu \quad \text{if and only if} \quad \int f(\alpha)\mu(d\alpha) \geq \int f(\alpha)\nu(d\alpha), \quad \forall f \in \mathcal{I}_m(\Lambda)$$

The natural generalization of a result on dominance in \mathfrak{R} is (see Torres[31]):

Theorem 8. $\mu \succeq \nu$ if and only if $\mu(A) \geq \nu(A)$ for every increasing measurable set A .

References

- [1] R.H. Rines *Create or Perish*, MIT Open Course Ware, (original publication, ca. 1964)
- [2] United States Patent and Trademark Office,
[<http://www.uspto.gov/web/offices/pac/doc/general/index.html>].
- [3] Japan Patent Office, [http://www.jpo.go.jp/quick_e/index_tokkyo.htm].
- [4] UK Patent Office, [<http://www.patent.gov.uk/patent/whatis/definition.htm>].
- [5] European Patent Office, [<http://www.european-patent-office.org>].
- [6] Canadian Intellectual Property Office,
[http://strategis.ic.gc.ca/sc_mrksv/cipo/patents/pat_gd_main-e.html].
- [7] Jaffe, Adam and Josh Lerner, *Innovation and Its Discontents*, Princeton University Press, 2004.
- [8] Boldrin, M. and D. K. Levine, *Against Intellectual Monopoly* at
[<http://www.dklevine.com/general/intellectual/against.htm>]
- [9] Alan Simpson, MP, Nicholas Hildyard and Sarah Sexton, “No Patents on Life A Briefing on the Proposed EU Directive on the Legal Protection of Biotechnological Inventions”
[<http://www.thecornerhouse.org.uk/item.shtml?x=51955>].
- [10] J. Bessen (2007), “Holdup and Licensing of Cumulative Innovations with Private Information”, mimeo, Research on Innovation.
- [11] J. Bessen and R. M. Hunt (2004), “An Empirical Look at Software Patents”, Working paper No. 03-17/R, Boston University, March 2004.
- [12] J. Bessen and E. Maskin (1999/2006) “Sequential Innovation, patents, and imitation”, Working paper No. 00-01, Massachusetts Institute of Technology, November 1999, revised March 2006.
- [13] Ted O’Donohue, Susan Scotchmer and Jacques-Francois Thisse, “Patent Breadth, Patent Life, and the Pace of Technological Progress”, *Journal of Economics and Management*, Vol 7, No. 1, pp 1-32, 1998.
- [14] Michael A. Heller and Rebecca S. Eisenberg, (1998), “Can Patents Deter Innovation? The Anticommons in Biomedical Research”, *Science* May 1998, Vol. 280. no. 5364, pp. 698 - 701.
- [15] Schumpeter, J. (1942), “Capitalism, Socialism and Democracy”, Allen and Unwin: London, 1976.

- [16] PUBLIC RECORD VERSION, In the Matter of Rambus, Incorporated, Docket No. 9302. OPINION OF THE COMMISSION, By HARBOUR, Commissioner, for a unanimous Commission.
- [17] John Markoff (2006), “In Silicon valley, A Man Without a Patent”,
[<http://www.nytimes.com/2006/04/16/technology/16wireless.html?ei=5...>].
- [18] Article: “Supreme Court Requests Brief From Solicitor General in Patent Obviousness Case” October 3, 2005, in [textittechlawjournal.com](http://www.textittechlawjournal.com) at:
[<http://www.techlawjournal.com/topstories/2005/20051003.asp>].
- [19] US Patent Law, [www.uspto.gov/web/offices/pac/mpep/consolidated_laws.pdf].
- [20] Quillen, Cecil D. Jr. (1993), “Proposal for the simplification and reform of the United States Patent System”, *American Intellectual Property Law Association Quarterly Journal*, Vol. 21, No. 3 (1993), pp. 189-212
- [21] Atkinson, S., A. Marco and J. L. Turner, (2006), “Uniformity and Forum Shopping in US Patent Litigation”, Mimeo, Terry School of Business, University of Georgia.
- [22] Henry, M. D. and J. L. Turner, (2005), “The Court of Appeals for the Federal Circuits Impact on Patent Litigation”, Mimeo, Terry School of Business, University of Georgia.
- [23] *The Economics of Patents and Copyright*, by Francois Lévésque and Yann Ménière at <http://www.cerna.ensmp.fr/CVs/Leveque.html>.
- [24] Torres, Ricard (1990), “Stochastic Dominance”, Discussion paper 905, Northwestern University, 1990.
- [25] “Motion for leave to file amicus curiae brief and brief of twenty-four intellectual property law professors as amici curiae in support of petitioner” in the case of KSR INTERNATIONAL CO., Petitioner, v. TELEFLEX INC. and TECHNOLOGY HOLDING CO., Respondents. No. 04-1350 in the Supreme Court of the United States.
- [26] SUPREME COURT OF THE UNITED STATES, Oral arguments in the case of KSR INTERNATIONAL CO., Petitioner v. TELEFLEX, INC., ET AL. No. 04-1350. Washington, D.C. Tuesday, November 28, 2006
- [27] Webbink, M. H. (2005), “A New Paradigm for Intellectual Property Rights in Software”, *Duke Law and Technology Review*, 0012.
- [28] Robert Thibadeau (2004), “Thomas Jefferson and Intellectual Property including Copyrights and Patents”, at [<http://rack1.ul.cs.cmu.edu/jefferson/>].

- [29] Zdnet magazine, “U.K. judge frowns on software patents”, January 13, 2006, at:
[http://news.com.com/U.K.+judge+frowns+on+software+patents/2100-1012_3-6027097.html?tag=st_1h].
- [30] Gilfillan, S.C. (1964), “Invention and the Patent System , Materials Relating to Continuing Studies of Technology, Economic Growth, and the Variability of Private Investment,” Joint Economic Committee, Congress of the United States, (Washington, D.C.: Government Printing Office).
- [31] Torres, Ricard. (1990), “Stochastic Dominance”, Northwestern Discussion paper.
- [32] Surowiecki, J. (2005), “Blackberry Picking”, *The New Yorker*, The Financial Page, December 19, 2005
- [33] *Innovation and Incentives*, by S. Scotchmer, publ. MIT press, 2004.