International Journal of Industrial Organization 3 (1985) 421–438. North-Holland

SOONER OR LATER?

Inventive Rivalry and Welfare

Lawrence M. DEBROCK*

University of Illinois, Champaign, IL 61020, USA

Robert T. MASSON*

Cornell University, Ithaca, NY 14853, USA

Final version received April 1985

An important branch of the literature on research and development expenditures has focused on the timing of innovation under different assumptions about the degree of inventive railway. This paper examines the timing decision under conditions of growing technological opportunities. Unlike earlier work, we explicitly treat social welfare in a second-best world where patent protection is used to spur innovation. We find that, given reasonable parameters, the orderings of private introduction times *vis-à-vis* the socially optimal time are ambiguous. We conclude that any welfare ranking would require a complete specification of all parameters (technical, behavioral and structural) surrounding each specific technological opportunity.

1. Introduction

In several recent studies economists have developed models with which to analyze firm invention and innovation. [For an excellent survey, see Kamien and Schwartz (1982).] Much of the theory has posited a competitive final product market and examined a process invention (i.e., a production cost savings) under varying assumptions concerning the number of rivals who are trying to patent the new invention. The market for innovation has generally been analyzed for a single inventor (the 'monopoly' inventor) and for numerous (potential) rivals (the competitive case).

One important branch of this literature focuses on the timing of innovation. There are two important decision variables which may be considered: (1) when to start the inventive process, and (2) how rapidly to proceed with this process (i.e., how soon will the process be 'successful'). Clearly, the earlier a firm achieves a successful innovation the sooner it will reap the rewards of

*We would like to acknowledge the helpful comments of Henry Wan, Robert Stoner, and two anonymous referees. This research was partially funded by a grant from the Bureau of Economic and Business Research, University of Illinois and National Science Foundation grant no. SES-8111237.

0167-7187/85/$3.30

innovation. Further, if there are potential rival innovators, the earlier the success, the greater the probability of achieving a blocking patent, preempting the market from others. Countering these gains from early success are potential costs of developing too rapidly. There may be growth in the potential market for the product or in the state of knowledge pertaining to the prospective invention in research and development labs across the economy. If this is true the net gains from R&D in a given industry will also be improving. Moving too quickly forecloses the firm from the benefits of such growth. Another cost of rapid development is associated with altering the development time. As a firm attempts to 'compress' the time between starting the inventive process and successful innovation, the total costs of an invention may rise.

The literature on the timing of innovation has generally considered one or the other problem: the optimal starting time or the optimal development speed. The former literature focuses on technological opportunities that grow over time while the latter assumes that at some point in time a profitable inventive opportunity becomes well-known and rivalry consists of racing to be the first firm to establish a patent.[1] In the present study we follow the first branch of the literature. We suppress the compression time question to examine the question of timing of invention under the assumption that technological opportunities are growing over time. Upon innovating the firm forecloses itself to further benefits flowing from future growth in technology.

Most previous studies have made the implicit or explicit assumption that the monopoly introduction time is a good proxy for the socially optimal introduction time. Barzel (1968), using an instantaneous preemptive starting time (no compression costs) model, found that rivalry always leads to 'premature' introduction. Examining the effects of rivalry in time-compression models, Scherer (1967) and Kamien and Schwartz (1972) demonstrate that competitive inventors/innovators can introduce either before or after the monopoly benchmark. In each of these studies, consumer surplus gains are not explicitly modeled. That is, the competitive introduction time is only compared to the non-rivalrous (monopoly) introduction time. Although these studies implicitly treat the monopoly introduction time as, in some sense, optimal, a monopolist is only certain to make the socially optimal decision if it reaps *all* of the benefits.[2]

Recently Dasgupta and Stiglitz (1980) and DeBrock (1980) made explicit welfare comparisons. The Dasgupta–Stiglitz model focuses on (1) a monopo-

[1]It should be noted that many authors allow innovators to choose (endogenously) invention size subject to an invention production function. Our point is that this production function, and its profitable inventive opportunities, are suddenly and exogenously put before the industry.

[2]Kamien and Schwartz (1972) explicitly note the possibility that the monopolist may not introduce at the socially optimal time, but they appeal to the earlier literature and use the monopoly time as the standard for comparison. Loury (1979) explicitly assumes infinite patent life and first degree price discrimination to make his monopoly time first-best.

list in both the invention and the product market, (2) competition in both the invention and the product market, and (3) invention by a new firm entering a monopolistic product market. They analyze these cases in a model of process innovation concerned with the issue of compression of development time. Their market incentive models are then compared with a 'socially managed' or first-best economy. Their first-best model derives a benchmark (socially optimal) introduction time by maximizing surplus under the assumption that the product market will have price at marginal cost before the process invention *and at the revised marginal cost* after the invention. They conclude that a monopolistic inventor (which also has a monopoly in the product market) would invent or innovate at a later time than a socially managed system, while a competitive system may innovate before or after a socially managed system.[3]

Our analysis extends DeBrock (1980). We examine the comparative statics of altering solely one thing, the degree of rivalry in the inventive process. Starting with competitive *product markets* both before and after invention, we then allow for a single inventor who has a 'monopoly' *for the invention/innovation* or numerous inventive rivals who are 'competitors' *for the invention/innovation.* Unlike Dasgupta and Stiglitz, we do not assume that the single 'monopolistic' inventor only occurs for monopoly product markets, nor do we assume that 'competitive' rival inventors only occur for competitive product markets. As was eloquently discussed in Nelson and Winter (1982), the important policy consideration for the 'Schumpeterian trade-off' is the market structure in innovation rather than the structure of the product market.

We examine innovation in a model of growing technological opportunities. In contrast to the Dasgupta–Stiglitz first-best approach, we assume that a patent system will be used to generate innovation. A patent system is an inherently second-best tool as it operates by creating a period of time in which price and marginal cost diverge. This is the prevalent mode of social inducement of innovative activity in our economy. We arrive at the general result that one cannot a priori determine the degree of innovative rivalry that

[3]The intuition is that a monopoly with constant returns and facing linear demand produces $x^m = x^c/2$ where x^c is the competitive output rate. If marginal cost were lowered from MC^0 to MC' the monopolist gains $(MC^0 - MC')x^m + (MC^0 - MC')\Delta x^m/2$ where the first term is cost savings on inframarginal units and the second term is the difference between marginal revenue and MC' over the range of new output, Δx^m. This is identically equal to one half the gain in a first-best society model where initial output is $x^c = 2x^m$ and the final output is $x^c + \Delta x^c = 2x^m + 2\Delta m$. The competitive gain if one firm can invest first and preempt all others is $x^c\ (MC^0 - MC')$. Hence the marginal value of introducing more rapidly is highest for the first-best case, slightly lower for the competitive case, and half of the first-best for the monopoly case. If the costs of compression are the same across groups the results are clear. Introducing uncertainty, e.g., that a firm in a competitive environment could try to race another firm in this environment, can yield greater expenditures on speed in a competitive environment.

is socially superior; the socially superior structure is not only parameter dependent, but highly sensitive to the parameter estimates.

The paper is organized as follows. The next section establishes the assumptions and structure of the model. Section 3 then describes the social properties of the equilibria under different degrees of innovative rivalry. Parametric simulations are used to demonstrate that the results are highly sensitive to parameter value changes within the zone of empirically reasonable parameter values.

2. The structure of the model

2.1. The approach to modelling

While it is straightforward to develop a growing frontier model in a very general form, complex sign restrictions result. To vastly simplify the analysis we use restrictive specific functional forms from the beginning. In justification, note that if we can demonstrate that the answers to our primary questions are parameter dependent in this simple model, they will remain so in a more general model.

We find this approach of using examples and counter-examples to be particularly pleasing because we can show complete generality by simply varying one or two parameters. In other words, our results indicate that the introduction times under different market structures may be ordered in any manner vis-à-vis the second-best social optimum. Furthermore, we can show that by simply varying one parameter we may shift from the conclusion that $W_m > W_c$ to $W_c > W_m$ where W_m and W_c are net social welfare in the monopoly/single inventor case and the competitive/rivalrous inventors case, respectively.

2.2. The basic assumptions

We consider the case of a patentable process innovation. The patented innovation is attained by the first firm to commit its R&D investment expenditures at a point in time.[4] The firm(s) will chose expenditures and timing to maximize their present values of profits. Recognizing its limitations, we employ the standard welfare measure, the present value of profits plus consumer surplus.[5] (We assume throughout that the social rate of discount is identical to the private rate, r.)

[4] By so doing, we suppress time compression considerations. One can imagine an innovation with a unique, deterministic development time. The qualitative results are unchanged.

[5] We also note that there are other effects of invention. Especially with patenting there may be numerous externalities. The invention must be published to be patented, leading to dissemination of knowledge and growth in other technological opportunities. Other externalities may arise, from unanticipated applications to related technology, to creation of negative pollutants. To the extent that the externalities are positive, this may tend to raise the social value of more rapid innovation. However, as the model demonstrates, more than the level of externality is important, its rate of growth over time is important as well.

The basic assumptions are:

(A.1) The *product market structure* is perfectly competitive prior to the process improvement and through the time the patent expires.

(A.2) The *patent system* grants a perfect monopoly over a particular production process improvement and all potential substitutes to this improvement for T years after the date of innovation, τ. Rights to use a patent may be licensed to other firms with no transactions costs and license fees may be set to extract the entire private surplus from the use of the patent.

(A.3) *Demand and cost* conditions for the final product market are posited to be of the following simple forms. Demand (except where otherwise noted) is characterized by a static, negatively sloped linear demand curve, $P=a-bq$. The cost side is, with the exception of the innovation, also static. It is represented by positive, horizontal marginal production cost curves.

(A.4) All inventions are of the '*run-of-the-mill*' variety. The run-of-the-mill invention [Nordhaus (1969) uses this term, Dasgupta and Stiglitz (1980) refer to it as a 'small invention'] assumption is a common starting point for analysis. This assumption is simply that the cost savings from a new invention are sufficiently small that the invention does not affect the quantity of final product supplied until the patent expires.

These initial assumptions are summarized graphically in fig. 1. At date τ a new invention is introduced which achieves a production cost savings of (MC^0-MC'). It is instantly licensed to all producers at a license fee per unit of final output, l, equal to its cost saving (minus an infinitesimal margin for

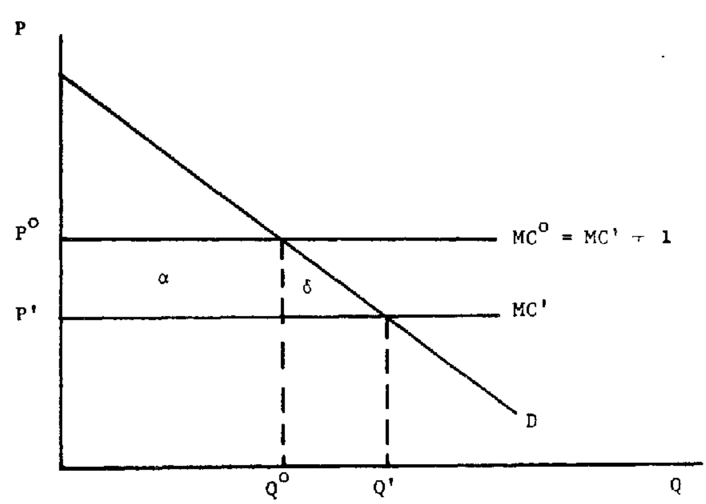


Fig. 1. Run-of-the-mill invention.

the licensee to be willing to use the new technology.) The competitive price before time τ is $P^0 = MC^0$ determined by technology and input prices. After time τ, but before the patent expires at time $\tau + T$, the competitive price remains at $P^0 = MC^0 = MC' + l$. There are no consumer surplus gains during this period. Clearly, output remains at Q^0 at least until time $\tau + T$, and the flow profits from innovating are equal to α until the patent expires. After the patent expires, l is no longer charged. Then price instantly falls to P', output expands to Q', and the flow of consumer surplus gains from the invention rises from zero to $\alpha + \delta$.

To establish a reference for analytical ease, we shall follow a normalizing convention of $P^0 = Q^0 = 1$. Hence under normalization $\alpha = 1 - P'$. Further, we note that for any linear demand curve, the value of α implies a unique δ; given linear demand, $\delta = \gamma\alpha^2$, where γ is half the value of the elasticity of demand.[6]

Finally, we must specify the technological opportunity set.

(A.5) *Inventive opportunities* are described by

$$\alpha = \alpha(I, \tau),$$

where $\alpha(I,0) - I < 0$, $\alpha_I > 0$, $\alpha_{II} < 0$, $\alpha_\tau > 0$, $\alpha_{\tau\tau} \leqq 0$, and $\alpha_{I\tau} = 0$ whenever $\alpha > 0$. Further, α is bounded, $\alpha < 1$, and both $\lim_{\tau \to 0} \alpha_\tau = \infty$ and $\lim_{\tau \to \infty} \alpha_\tau = 0$.

The assumption $\alpha_\tau > 0$ means that information externalities are making it progressively easier to achieve any level of cost savings. These externalities can originate from many possible sources, from academic research to innovations in related areas. For example, on-going research in computers and high-tech electronics has had demonstrated effects on the efforts of research and development labs in unrelated industries across the economy.[7] Our thrust is to model this as growth in the benefits from inventive efforts as time passes. The assumption that $\alpha_{\tau\tau} \leqq 0$ is made in part for analytical convenience. It restricts the formulation to those cases in which informational externalities make any level of invention easier to achieve, but the benefits cannot accrue at an increasing rate.

The simplifying assumption that $\alpha_{I\tau} = 0$ for all $\alpha > 0$ is not required but it provides many advantages: (1) it is sufficient for showing heretofore not demonstrated generality of the results on timing of innovation, (2) it

[6]Recall $P = a - bQ$. Note that $\delta = (P^0 - P')(Q' - Q^0)/2$ and that $(P^0 - P') = b(Q' - Q^0)$ by the linearity of demand and $\alpha = (P^0 - P')$ by normalization. It follows that $\delta = (\alpha^2/2b)$. Note that b is the reciprocal of the elasticity of demand at P^0, Q^0.

[7]The use of micro computer chips to regulate automobile functions was widely predicted several years before such a development became economically feasible. Likewise, manufacturers now anticipate breakthroughs in other fields which will facilitate their move toward automation.

considerably reduces analytical problems, and (3) it is a priori a reasonable case.[8] In effect, when $\alpha_{I\tau}=0$ the optimal level of investment, I, remains invariant to the timing of innovation τ. Thus with $I_c=I_m$, we may work with a single value of τ_s and meaningfully examine $\tau_s \gtreqless \tau_c$ and $\tau_s \gtreqless \tau_m$. Because $I_c=I_m$ the only difference which can affect welfare is timing. For many cases this allows us to infer the structure of inventive rivalry that is socially dominant by simple examination of the orderings of the introduction times. Finally, the boundary restriction ($\alpha<1$) comes from the normalization and the limiting conditions on α_τ are simply Inada conditions used to eliminate generally trivial corner solutions from the analysis.

These assumptions can be used to provide structure to our model. The present discounted value of the innovation, V, is a function of both the patent life, T, and the introduction time, τ. This private value at the planning time (time zero) is

$$V=\int_{\tau}^{\tau+T} \alpha(I,\tau)e^{-rt}dt-Ie^{-r\tau}. \tag{1}$$

After patent expiration, consumer surplus benefits accrue. Recalling that $\delta(I,\tau)=\gamma\alpha(I,\tau)^2$, the present discounted value of such benefits can be written as

$$C=\int_{\tau+T}^{\infty} [\alpha(I,\tau)+\gamma\alpha(I,\tau)^2]e^{-rt}dt. \tag{2}$$

Social welfare is measured as net surplus, the sum of firm and consumer benefits. Its discounted form is

$$W=\int_{\tau}^{\tau+T} \alpha(I,\tau)e^{-rt}dt+\int_{\tau+T}^{\infty} [\alpha(I,\tau)+\gamma\alpha(I,\tau)^2]e^{-rt}dt-Ie^{-r\tau}. \tag{3}$$

All of these equations represent discounted values at the planning time, period zero. In order to better interpret later results it is also useful to define

[8]That $\alpha_{I\tau}=0$ is an a priori reasonable case should be clear. For instance, for a firm thinking of developing a new gasoline production process we could envision three possible scenarios. We could assume that at time τ^0 the firm decided to spend I^0 dollars on a process to save production costs of X^0 cents per gallon. However, before it starts its R&D process it learns of a breakthrough in a related field of chemistry. It is possible that: (a) the optimal decision is to achieve approximately an X^0 cent per gallon cost savings but with considerably lower investment ($I'<I^0$), or (b) the optimal decision is to reduce costs per gallon by an amount far in excess of X^0 cents per gallon by spending more than originally budgeted, ($I'>I^0$). Since both cases are plausible it is also plausible that the increased knowledge could lead to an optimal decision in which $I'=I^0$.

terms which represent the discounted values of each surplus *at the time of introduction*, time τ.

Noting $\int_0^\infty e^{-rt}dt=(1/r)$ and defining $\phi=(1-e^{-rT})$, these are as follows:

$$v=\int_0^T \alpha(I,\tau)e^{-rt}dt-I=\alpha(I,\tau)\frac{\phi}{r}-I, \tag{1'}$$

$$c=\int_T^\infty [\alpha(I,\tau)+\gamma\alpha(I,\tau)^2]e^{-rt}dt=[\alpha(I,\tau)+\gamma\alpha(I,\tau)^2](1-\phi)/r, \tag{2'}$$

$$w=\int_0^T \alpha(I,\tau)e^{-rt}dt+\int_T^\infty [\alpha(I,\tau)+\gamma\alpha(I,\tau)^2]e^{-rt}dt-I \tag{3'}$$

$$=\alpha(I,\tau)\frac{\phi}{r}+[\alpha(I,\tau)+\gamma\alpha(I,\tau)^2](1-\phi)/r-I.$$

Thus, we can now represent the private and social values at time zero as

$$V=ve^{-r\tau}, \tag{1''}$$

$$C=ce^{-r\tau}, \tag{2''}$$

$$W=(v+c)e^{-r\tau}=we^{-r\tau}. \tag{3''}$$

With this model structure we can now examine the issue of optimal timing of invention.

3. Inventive rivalry and second-best social optima

We first derive the analytic conditions for firm and social optima. After demonstrating these welfare results we present a simulation to demonstrate the sensitivity of the results to changes in parameters (demand elasticity, growth, etc.).

3.1. Optimal invention

Consider first the single inventor case: the 'monopoly inventor'. The monopoly inventor's first-order conditions to maximize present value $V(I,\tau)$ from eq. (1'') are

$$V_I=(\alpha_I\phi/r-1)e^{-r\tau}=0, \tag{1'''}$$

$$V_\tau=\left(\alpha_\tau\phi/r-r\left(\alpha\frac{\phi}{r}-I\right)\right)e^{-r\tau}=0. \tag{1''''}$$

As r and ϕ are independent of τ, the assumption $\alpha_{I\tau}=0$ permits (1‴) to be solved for a level of $I=I^*$ which is time invariant. This level is illustrated in fig. 2. Note how, as introduction time is delayed, the frontier shifts vertically.[9] Recognizing that $I=I^*$, (1⁗) can be solved for the remaining unknown, τ. The monopoly inventor's optimal invention time thus solved for is denoted τ_m.

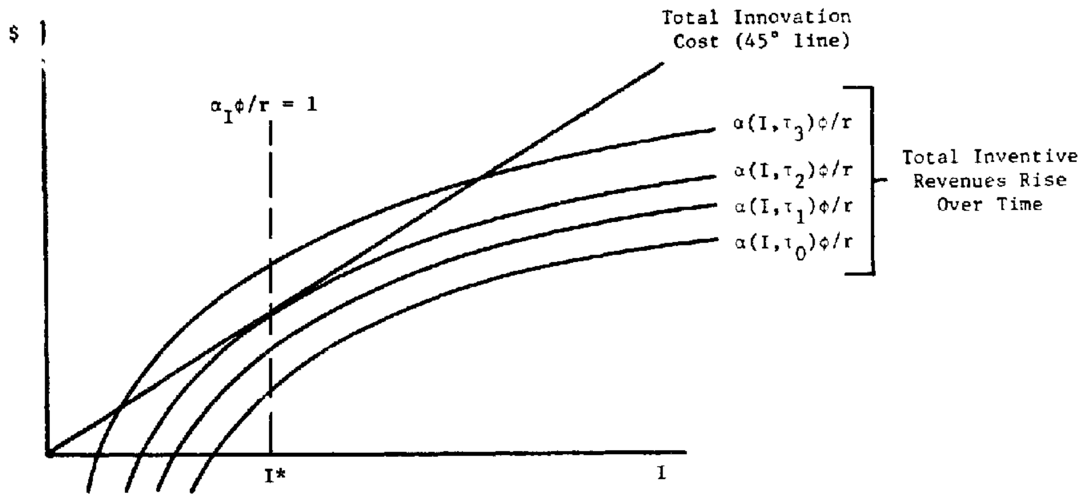


Fig. 2. Inventive opportunities.

To better understand the solution for τ_m it is useful to rewrite the condition implied by (1⁗) in terms of the value function which gives firm value at the time of invention: $v(I,\tau)$. By dividing by $e^{-r\tau}$, (1⁗) becomes

$$v_\tau - rv = 0. \qquad (1''''')$$

The interpretation is now readily apparent. By delaying invention the firm gains v_τ from the growth of the frontier, but delays its realization. The cost of delay is the foregone interest on v of rv. Invention occurs when v_τ has converged upon rv from above as at point B in fig. 3. This illustrates the solution for τ_m.

The competitive rivalry, multiple (potential) inventor case in this model contrasts strikingly with that in the time-compression models such as Scherer (1967), Kamien and Schwartz (1972), or Dasgupta and Stiglitz (1980). In their models, a profitable innovation possibility suddenly appears. When the

[9]The continuation of the curve in the negative quadrant can be interpreted as the vertical displacement of some further $\alpha\phi/r$ projected back to time τ. To see this note that the assumption $\alpha_{I\tau}=0$ for $\alpha>0$ assures that if $\alpha_\tau>0$ at τ_c then for $\tau>\tau_c$ we will see an $\alpha>0$ for some $I<\bar{I}$. In the positive quadrant the slope of $\alpha\phi/r$ is time invariant for any I.

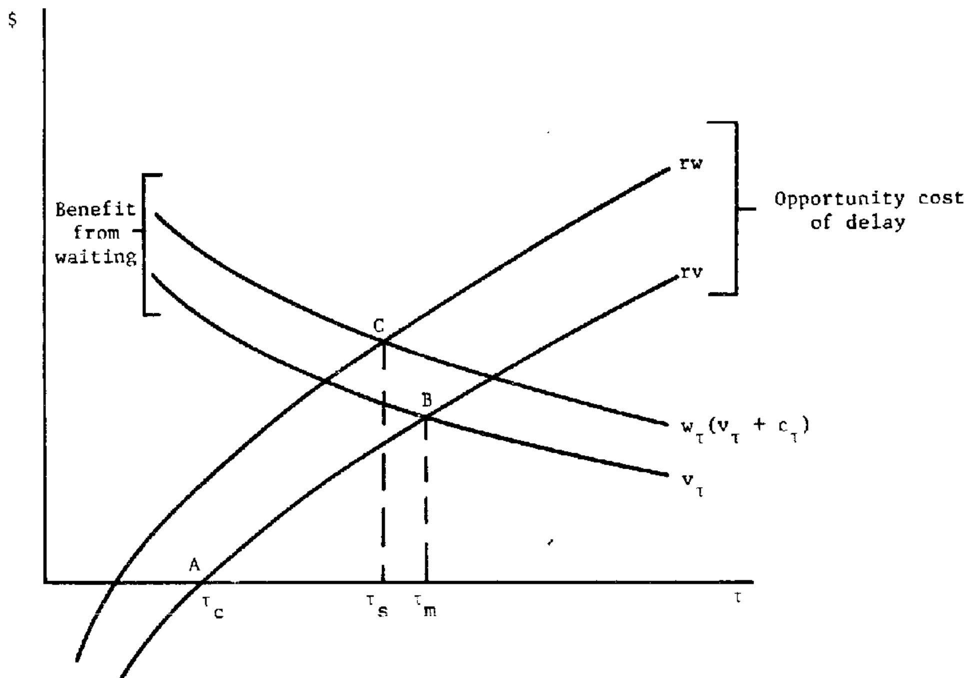


Fig. 3. Optimal timing.

new technological frontier is announced, some number of inventors, determined by a zero profit/value condition, will 'race' to gain the first patent. By contrast, our model assumes a continuously growing frontier. With Nash behavior of potential entrants we arrive at a case in which only one firm will invest in invention even in the competitive case. (This is as in Barzel's or Dasgupta–Stiglitz's 'certainty case'.) Each of numerous firms will consider the potential profits/value from moving to invent. As $v(I^*,\tau)=[\alpha(I^*,\tau)\phi/r-I^*]$ becomes non-negative, then any firm would find I^* profitable if there were no other firms investing in I and unprofitable otherwise. Hence, when $V(I^*,\tau)=v(I^*,\tau)e^{-r\tau}$ becomes non-negative the first firm to invest preempts invention for all other firms. The maximizing conditions for the winner are (i) an investment first order condition identical to (1'''), and (ii) a zero-profit (value) preemption condition. These are, respectively,

$$V_I=(\alpha_I\phi/r-1)e^{-r\tau}=0, \tag{4a}$$

$$V=(\alpha\phi/r-I)e^{-r\tau}=0. \tag{4b}$$

As before, (4a) determines the identical $I=I^*$ and then (4b) can be solved for $\tau=\tau_c$, the competitive rivalry invention time. (This is time τ_2 in fig. 2, where the frontier is tangent to the cost line, and is labeled τ_c in fig. 3.) Contrasting (4b) and (1''''), recognizing $I=I^*$ in both cases, $\tau_m>\tau_c$ (the monopoly inventor delays longer before invention) unless $\alpha_\tau(I^*,\tau_c)=0$. The monopoly inventor never invents prior to the competitive inventor in our instantaneous preemptive invention model because to do so would yield a negative present discounted value.[10]

With knowledge of these conditions, our remaining task is to find the second-best (i.e., given a patent system) socially optimal innovation time, τ_s, and contrast it with τ_m and τ_c. It is at this point that the assumption of $\alpha_{I\tau}=0$ for $\alpha>0$ plays its primary simplifying role. As noted above, $\alpha_{I\tau}=0$ fixes $I=I^*$ for both the competitive and the monopoly cases. The socially optimal introduction time *given a patent system* can then be derived from maximizing (3″) with respect to timing. After some manipulation, this first-order condition is represented in (3‴).

$$W_\tau=\{\{\alpha_\tau\phi/r+[\alpha_\tau+2\gamma\alpha\alpha_\tau](1-\phi)/r\}$$
$$-r\{[\alpha\phi/r-I]+[\alpha+\gamma\alpha^2](1-\phi)/r\}\}\,\mathrm{e}^{-r\tau}=0. \qquad (3''')$$

In (3‴) the consumer surplus effects are separated from the producer surplus effects and the gains from waiting are separated from the costs of delay. In terms of values derived at the invention date the terms in (3‴) are, respectively,

$$W_\tau=\{(v_\tau+c_\tau)-r(v+c)\}\mathrm{e}^{-r\tau}=0, \quad \text{or}$$
$$=\{w_\tau-rw\}\mathrm{e}^{-r\tau}=0. \qquad (3'''')$$

Again the solution is one at which the gains from waiting, w_τ, have fallen to the opportunity cost of foregone welfare from delay, rw. An example is illustrated in fig. 3. At point C, $w_\tau=rw$, indicating the second-best socially optimal invention time, τ_s. For this example, $\tau_c<\tau_s<\tau_m$. It is illustrative to investigate the determinants of this ordering.

The second-best socially optimal timing will pre-date the competitive timing if and only if $w_\tau-rw<0$ at τ_c (i.e., at time τ_c the benefits from waiting have already fallen below the costs of foregone revenues from delay). More generally, by substituting the τ_c defined by (4b) into the w_τ from (3‴), this condition is

$$\tau_s \lesseqgtr \tau_c \quad \text{if and only if at } \tau_c\text{:}$$

$$\alpha_\tau\phi/r+(\alpha_\tau+2\gamma\alpha\alpha_\tau)(1-\phi)/r-r(\alpha+\gamma\alpha^2)(1-\phi)/r \lesseqgtr 0, \quad \text{or} \qquad (5)$$

$$v_\tau(I^*,\tau_c)+c_\tau(I^*,\tau_c)-rc(I^*,\tau_c) \lesseqgtr 0. \qquad (5')$$

[10] Kamien and Schwartz (1972) and Scherer (1967) found $\tau_c>\tau_m$ for some cases of extreme rivalry. However, their models concentrated upon the time compression aspects of introduction time, as discussed at the outset.

The interpretation of this condition is as follows: if at τ_c the gains in producer and consumer surplus from delaying introduction fall below the opportunity cost of consumer surplus foregone by delay, there would be a social improvement if the invention were to be introduced earlier (i.e., $\tau_s<\tau_c$). To compare τ_s and τ_m, one may substitute (1'''') in w_τ. This gives

$$\tau_s \gtreqless \tau_m \quad \text{if and only if at } \tau_m\text{:}$$

$$(\alpha_\tau+2\gamma\alpha\alpha_\tau)(1-\phi)/r - r(\alpha+\gamma\alpha^2)(1-\phi)/r \gtreqless 0, \quad \text{or} \tag{6}$$

$$c_\tau(I^*,\tau_m) - rc(I^*,\tau_m) \gtreqless 0. \tag{6'}$$

Since producer surplus has already been optimized, the question is now whether the gain in consumer surplus exceeds its opportunity cost foregone by delay.

Sign restrictions on (5), (5'), (6) and (6') are not readily apparent. This is not surprising as we shall prove, by example, that it is possible to have $\tau_c<\tau_m<\tau_s$, $\tau_c<\tau_s<\tau_m$, or $\tau_s<\tau_c<\tau_m$. For these cases we note that $W_m>W_c$, $W_m \gtreqless W_c$, and $W_m<W_c$, respectively.

3.2. Some intuitive interpretations

Optimization involves comparing the gains from waiting against the opportunity costs of delaying the payoffs. When there is inventive rivalry, the competing inventors weigh the potential revenues from invention against the cost of investment, I^*. When such profits become positive, they invest immediately in order to avoid being preempted. In the case without rivalry, the monopoly inventor simply compares the growth of profit opportunities against profits foregone by delay. In neither case does a firm consider consumer surplus. Clearly, the social optimal timing must take into account both the level of consumer surplus and its growth.

It is simple to construct heuristic examples demonstrating why τ_s may be greater or less than τ_c. These examples can yield some insights into what factors cause τ_s to vary relative to τ_c and/or τ_m. Consider a simple case in which $\tau_s<\tau_c<\tau_m$. Suppose that $\alpha_{\tau\tau}<0$ and that $\alpha(I^*,\tau)\phi/r$ rises over time, but only asymptotically approaches $I^*+\varepsilon$. This can be visualized in two ways. First, the $\alpha(\cdot)\phi/r$ curves in fig. 2 approach (but do not cross) a line ε above the total innovation costs curve or second, the rv curve in fig. 3 rises smoothly and is asymptotic to $(r\varepsilon)\to 0^+$. By assuming that ε is as small as we wish we can move τ_c (and τ_m) to values as large as we wish. But note that as $\varepsilon\to 0$, long before private value is positive the social value will be positive and finite. (E.g., note that $rw>rv$ in fig. 3, as $w-v=c>0$.) Also note that the social gains from delay after some date will be infinitesimal (both c_τ and v_τ

go to zero as v becomes asymptotic to ε so w_τ goes to zero while w is finite). Hence, as $\varepsilon \to 0$ we see finite τ_s while τ_c and τ_m both approach infinity: both competitive and monopolistic introduction will be delayed past the socially optimal time.

This case is directly analogous to the 'social overhead capital' problem. In a static model the social value of a project may be positive even with demand strictly below average cost. This occurs because the consumer surplus value of the infra-marginal units sold cannot be privately captured.

Next consider a case in which $\tau_s > \tau_c$. Suppose that $\alpha\phi/r$ is rising rapidly at τ_c. If this rise is sufficiently rapid then w_τ $(=v_\tau + c_\tau)$ will be so positive as to exceed rw $(=r(v+c)=rc)$ at time τ_c. Clearly, rc may be a very small finite number, and a sufficiently large value of v_τ can achieve $w_\tau > rw = rc$ at τ_c. In other words, if the rate of growth of social value is large enough, the social gains from waiting will exceed the costs of delay given any social value at τ_c.

These heuristic examples demonstrate some of the factors that determine whether τ_s is greater or less than τ_c. To construct these examples we simply varied the growth profile of $v(I^*, \tau)$. For this reason growth will be a major focus as we analyze simulated comparative statics in a parameterized version of the model below.

The other crucial variable we will focus on is demand elasticity. Although the variance of τ_s around τ_c can be discussed without considering demand elasticity, demand elasticity is crucial when considering the position of τ_s in the ranking. This follows directly from the role of demand elasticity in the size of consumer surplus. From (6′), $\tau_s > \tau_m$ simply means that $c_\tau > rc$ at τ_m as $v_\tau = rv$ at τ_m. But if demand elasticity is zero then $\gamma = 0$. If $\gamma = 0$ then $c_\tau = v_\tau(1-\phi)/\phi$ and $c = (v+I)(1-\phi)/\phi$. Thus, when demand elasticity is zero and $v_\tau = rv$, we have $c_\tau < rc$ and $\tau_s < \tau_m$. When demand elasticity is non-zero, however, c_τ can be greater than rc since c_τ rises geometrically in α, while v_τ is linear in α. To analyze this more fully we turn to a parameterized $\alpha(\cdot)$ function and simulate comparative statics.

3.3. Simulation and comparative statics

Simulation requires a functional form for $\alpha(\cdot)$. We use a simple additively separable specification of

$$\alpha = AI^\beta + g\tau - K. \tag{7}$$

With $\beta < 1$, this form exhibits diminishing marginal revenue product of I, constant absolute growth of α over time (g), and is initialized in time by setting the vertical shift term, K. The form in (7) does not assure that $W(\cdot)$ and $V(\cdot)$ are both concave or that α is bounded by 1. However, the function chosen has a single critical point and satisfies the second-order coditions at this point. The parameter values used in the simulation were selected to assure $0<\alpha<1$ at the critical point.

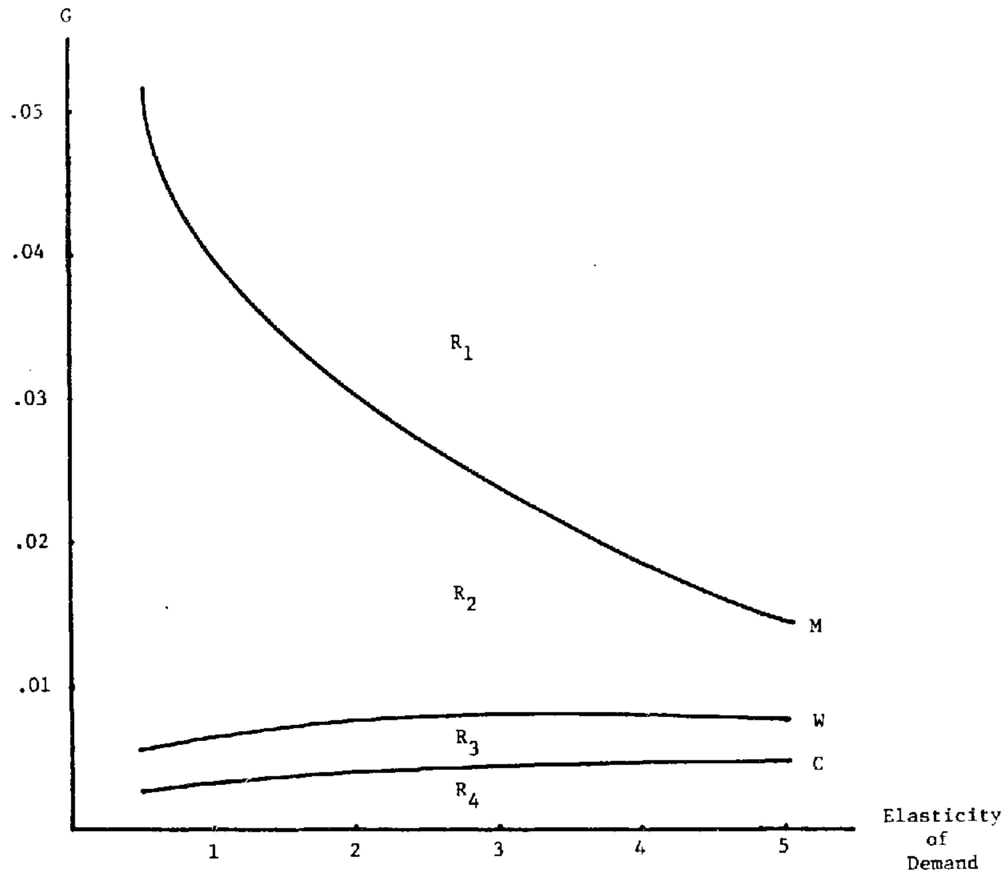
For the examples which follow it is necessary to choose values for the various parameters. We assume $A=1$, the opportunity cost of capital (r) is 10 percent, and, in light of empirical estimates by Mansfield, the parameter β is set to 0.10.[11] Patent life (T) is 17 years as in the United States. The shift parameter, K, serves only to initialize the time when the invention first becomes viable (τ_c). Since our focus is on relative timing, the qualitative results are insensitive to the selection of K under the condition $V(I^*,0)\leqq 0$.

Fig. 4 depicts the results of simulations when the two remaining parameter values, demand elasticity and (the absolute) technical growth, are varied. Recall that the private introduction time, regardless of the degree of inventive rivalry, is affected by the technological growth parameter but *not* by the demand elasticity in the final product market. However, the socially optimal introduction time is directly affected by this elasticity parameter. Thus, for any given demand elasticity it is possible to derive a value of the technological growth parameter, g, which would lead private firms to choose an introduction time equal to the socially optimal introduction time. The two curves, M and C, represent 'critical values' of demand elasticity and technological growth for the monopoly and competitive cases of inventive rivalry. For example, along the M locus, a monopolist would choose an introduction time identical to the (second-best) socially optimal time.

It follows that any combination of demand elasticity and technological growth which lies below the locus C would lead a competitively structured invention market to an introduction date later than the socially optimal time. Obviously, since $\tau_m>\tau_c$, a monopoly inventor would also lead to socially late introduction. Any point above M in fig. 4 would result in monopoly, and therefore competitive, introduction times earlier than the socially optimum introduction time.

It is easy to make welfare rankings for any combination of demand elasticity and growth which lies above the M locus ($W_m>W_c$) or below the C locus ($W_m<W_c$). For these cases the clear ordering of privately optimal introduction times relative to the second-best socially optimal provides an unambiguous guide. However, the interior between M and C requires additional analysis. To make welfare rankings in this region one must calculate the net surplus that would accrue under the relevant degree of rivalry. For each level of demand elasticity there exists a unique growth level at which the net surplus generated under either market form is identical (even though the introduction times differ, of course). This locus is labelled

[11]Mansfield (1965) developed empirical estimates of β, the 'returns to scale' of R&D. His estimate of β is 0.1, the value we use in our calculations. Simulation shows the orderings of introduction times to be sensitive to the assumed level of β in the zone of 0.1. Given potential estimation biases and differing β's across invention types, this sensitivity is worth noting.



Region	Timing		
R_1	$\tau_c<\tau_m<\tau_s$	so	$W_m>W_c$
R_2	$\tau_c<\tau_s<\tau_m$	and	$W_m>W_c$
R_3	$\tau_c<\tau_s<\tau_m$	and	$W_m<W_c$
R_4	$\tau_s<\tau_c<\tau_m$	so	$W_m<W_c$

Fig. 4. Welfare and timing as related to demand elasticity and growth.

W in fig. 4. Any combination of growth and demand elasticity above the line implies that society would favor the single monopoly inventor, and below the line it would prefer to have rivalrous/competitive inventors.

Interpretation of these results with respect to equilibrium growth rates, rather than simply growth, can be useful. For this reason several equilibrium growth rates of the gross returns function $\alpha(\cdot)$ are illustrated in table 1.[12]

The heuristic examples in section 3.2 involved the $v(\cdot)$ function either becoming asymptotic to some low value, or raising rapidly at first and then

[12] Tables of critical values for various parameters values in either type of market are available upon request from the authors. Sensitivity analysis of the various parameters indicated that the general shapes of the *M*, *C* and *W* loci were quite robust.

Table 1
Equilibrium growth rates at points on the M, C and W curves.

	Demand elasticity		
	0.5	1	5
M	0.0866	0.0822	0.0693
C	0.0185	0.0188	0.0209
W	0.0323	0.0317	0.0277

flattening. Despite the fact that the simulation is based upon linear demand and on α function (7) which is linear in τ, the intuition from above remains robust. For example, consider a scenario with a growth rate so low that $\tau_s < \tau_c$. The fact that at $v=0$ one has $w=c>0$ coupled with the low growth rates (and hence low $w_\tau = v_\tau + c_\tau$) implies an optimal introduction time $\tau_s < \tau_c$. As demand elasticity is raised, c_τ is raised by less than rc in the zone of $\tau_s = \tau_c$. (Note that given the geometric increases in c as α rises, c_τ is relatively low at 'early' time periods, like τ_c, but high during later time periods, like τ_m.) This causes the locus C, at which $\tau_s = \tau_c$, to be rising in elasticity.[13]

While the preceding analysis concentrated upon the timing relationships as they are affected by elasticity and growth, it should be clear that the loci M, C and W are also sensitive to the other parameters. For any given demand elasticity one may calculate the elasticity of these loci M and C with respect to the remaining parameters: r, β and T. These elasticities are presented at three levels of demand elasticity in table 2. All of the loci (especially C) are highly sensitive to parameter specification.

Table 2
Elasticities of g values underlying M, C and W with respect to r, β and T.

	Demand elasticity								
	0.5			1			5		
	M	C	W	M	C	W	M	C	W
r	+1.005	−0.711	+7.204	+1.006	−0.696	+7.127	+1.009	−0.607	+6.661
β	−0.069	+0.016	−0.169	−0.090	+0.032	−0.187	−0.129	+0.125	−0.297
T	−0.003	−1.67	−1.23	−0.004	−1.66	0.160	0.005	−1.55	−0.046

[13]The intuition from the heuristic example based upon high growth rates leading to later τ_s relative to τ_c is also apparent. Further, once time has progressed to τ_m, c_τ becomes larger due to its geometric relationship with α. In the zone of τ_m, higher elasticity is associated with $c_\tau + v_\tau$ being raised more than $r(c+v)$, causing the locus M at which $\tau_m = \tau_s$ to be declining in elasticity.

3.4. Summary and implications

We have demonstrated results which show that in our model $\tau_c < \tau_m$. We have further demonstrated that τ_s may precede or follow either τ_c or τ_m. In welfare terms we can summarize these results as

$$W_m > W_c \quad \text{if} \quad \tau_s > \tau_m > \tau_c,$$

$$W_m \gtreqless W_c \quad \text{if} \quad \tau_m \geqq \tau_s \geqq \tau_c,$$

$$W_m < W_c \quad \text{if} \quad \tau_m > \tau_c > \tau_s.$$

The answer to the question of whether inventive rivalry raises or lowers welfare is shown to be ambiguous even in a very simple model. Even though the answer can be made unambiguous by the use of parameter estimates, the results were shown to be highly sensitive to the parameter estimates. Somewhat more general models can lead to yet further ambiguity. For one example, Scherer (1967) and Kamien and Schwartz (1972) present models in which $\tau_c > \tau_m$ in some equilibria. Further, to the extent that timing of innovation raises the $\alpha(\cdot)$ frontier for yet another innovation, this information externality leads toward a preference, at the margin, for earlier introduction. Finally, note that when successful invention is not a foregone conclusion, rivalry may lead to independent paths towards the same goal, raising the probability of successful invention.

4. Conclusion

Nelson and Winter state that the crucial question for policy analysis of the 'Schumpeterian trade-off' is whether competitive rivalry raises or lowers welfare relative to the case of a single inventor. We have demonstrated that this question is extremely hard to answer. If we are to take some sort of patent system (or other system which does not legislate both price equal to marginal cost and optimal R&D investment behavior) as the basic environment within which R&D takes place, ambiguity is often the result. For a very simple formulation of the problem we cannot predict, in the absence of accurate empirical estimation of industry specific parameters, whether rivalry or monopoly is socially preferred for innovation. Further, we cannot predict whether an additional costless incentive system designed to either speed or slow R&D would be welfare enhancing or welfare detracting simply on the basis of the degree of inventive rivalry.

The intuitive reasons for this should be clear from the text. Competitive rivalry forces the inventor to ignore the rate of growth of benefits over time. Both a monopoly inventor and a competitive inventor will ignore the consumer surplus triangle (which grows geometrically). However, both the rate of growth of benefits and the triangle losses are crucial to the (second-best) social optimum. In addition, the present values of these terms are quite sensitive to demand growth, demand elasticity, returns to scale, interest rates and patent life.

Given the level of abstraction in the model we can only conclude with the following note. Various alternative formulations could be used for this problem. However, the fundamental ambiguity we have demonstrated here implies that the answers to the policy questions noted above are empirical in nature. Unfortunately, the primary questions will be difficult to resolve empirically. First, each possible innovation scenario (e.g., new product, compression cost) must be completely analyzed in theory with all externalities accounted for (e.g., the value of mandatory publication of patent information). Then each case must be empirically described, a vast undertaking given the unobserved nature of much of the relevant information. The next step might be to assess the empirical incidence of each case, and then aggregate individual welfare measures into some weighted average welfare sum. Finally, it would still be necessary to forecast the future incidences of cases and devise that corresponding weighing scheme. Only then can there be a basis for generally advocating more or less rapid innovation or more or less competitition in the innovation process. This gloomy result does not deny occasionally being able to analyze ex post a single invention and/or innovation, but it shows that generality may only come in the distant future, if ever.

References

Barzel, Y., 1968, Optimal timing of innovation, Review of Economics and Statistics 50, Aug., 348–355.

Dasgupta, P. and J. Stiglitz, 1980, Uncertainty, industrial structure, and the speed of R&D, Bell Journal of Economics 11, Spring, 1–28.

DeBrock, L.M., 1980, Inventive rivalry and social efficiency, unpublished dissertation (Cornell University, Ithaca, NY).

Fethke, G.C. and J.J. Birch, 1982, Rivalry and the timing of innovation, Bell Journal of Economics 13, Spring, 272–279.

Futia, C., 1980, Schumpeterian competition, Quarterly Journal of Economics 48, June, 675–695.

Kamien, M. and N. Schwartz, 1972, Timing of innovations under rivalry, Econometrica 40, Jan, 43–60.

Kamien, M. and N. Schwartz, 1982, Market structure and innovation (Cambridge University Press, Cambridge).

Loury, G.C., 1979, Market structure and innovation, Quarterly Journal of Economics 93, Aug., 395–410.

Mansfield, E., 1965, Rates of return from industrial research and development, American Economic Review 55, May, 310–322.

Nelson, R.R. and S.G. Winter, 1982, The Schumpeterian tradeoff revisited, American Economic Review 72, March, 114–132.

Nordhaus, W.D., 1969, Invention, growth, and welfare (MIT Press, Cambridge, MA).

Scherer, F.M., 1967, Research and development resource allocation under rivalry, Quarterly Journal of Economics 81, Aug., pp. 359–394.

Scherer, F.M., 1980, Industrial market structure and economic performance, 2nd ed. (Rand McNally, Chicago, IL).