Review of Economic Studies (1997) 64, 265-288 0034-6527/97/00130265$02.00

Waves of Creative Destruction: Firm-Specific Learning-by-Doing and the Dynamics of Innovation

JEREMY C. STEIN
MIT Sloan School of Management and NBER

First version received June 1995; *final version accepted September* 1996 (*Eds.*)

This paper develops a model of repeated innovation with knowledge spillovers. The model's novel feature is that firms compete on *two* dimensions: (1) product quality, where one firm's innovation ultimately spills over to other firms; and (2) distribution costs, where there are no spillovers across firms and where learning-by-doing on the part of incumbent firms gives them a competitive advantage over would-be entrants. Such firm-specific learning-by-doing has two important consequences: (1) it can in some circumstances dramatically reduce the long-run average level of innovation; (2) it leads to endogeneous bunching, or waves, in innovative activity.

1. INTRODUCTION

Recent work in the literature on innovation and growth has emphasized the process that Schumpeter (1942) labeled "creative destruction." In the models of Aghion and Howitt (1992), Grossman and Helpman (1991*a*, *b*), Segerstrom (1991), Segerstrom, Anant and Dinopoulos (1990), and Caballero and Jaffee (1993), new, higher-quality products are introduced by new firms, thereby displacing incumbent firms. The introduction of these new products in turn sets the stage for yet another round of innovation, entry, and displacement, because of the existence of knowledge spillovers—once a new product is introduced, future generations of innovators can learn from it and improve upon it.

The image that comes through in these papers is of a very fluid corporate sector, with new companies continually pushing aside existing ones. Indeed, this fluidity is seen as an essential element in the growth process: it is only through the destruction of existing firms' market shares and profits that new, better products—and the embedded knowledge that accompanies them—come into being. Yet casual empiricism suggests that the corporate landscape may be much less fluid than these models envisage. Many companies maintain their market shares in given product areas for a long time. Often they manage to do so even as the technological environment in which they operate changes dramatically.

These are two broad reasons why established firms might be expected to survive in a rapidly changing environment. First, and most simply, established firms may just have a comparative advantage in innovation. This would imply that established firms naturally tend to stay "on the cutting edge" of new products and new technology, since they actually pioneer the majority of improvements. If this is the case, the survival of such firms would be no surprise—they survive because they are either the highest quality, or most efficient producers at any point in time.[1]

1. The literature has identified a number of factors that may help to foster innovation inside already-established firms. Gilbert and Newbery (1982) argue that incumbents are likely to have a greater strategic incentive to invest in innovative activity than potential entrants when the innovations in question are incremental

A second possibility is that established firms are *not* always on the cutting edge of new products and technology, but rather, they have a number of *other* competitive weapons at their disposal that allow them to fend off more innovative newcomers and to thereby forestall the process of creative destruction. For example, long-established firms are likely to have a base of loyal customers and well-developed distribution networks that allow them to market whatever products they do have more effectively. This latter explanation of corporate survival in the face of outside product innovation fits well with the literature on competitive strategy (e.g., Porter (1980)). Work in this area repeatedly emphasizes that innovative, high-quality products are but one route to "competitive advantage", and that loyal customers and a strong distribution network can also be crucial links in the "value chain". The customer base story also seems to coincide with how some managers of established firms described their own competitive strengths. For example, Frank Perna, CEO of Magna Tek, a $1.3 billion manufacturer of electrical equipment, argues that: "I think the core competence of our company is really its distribution channels—*far more so than its products.*" (emphasis added)[2]

In general terms, what is significant about customer bases in this context is that they represent the accumulation of learning-by-doing that is *both firm- and product-specific.* The firm-specificity implies that potential entrants cannot take advantage of the knowledge about customers that has been created by established firms. And conversely, the product-specificity implies that established firms cannot simply knock off or acquire any new products developed by potential entrants, and sell them at lower cost through their existing distribution channels. Thus both established firms and innovative newcomers have their own distinct, non-transferable sources of advantage, which allows a meaningful tension to arise. Although I will use the euphemism "customer bases" as shorthand throughout, it should be clear that what I have in mind is this broader notion of firm- and product-specific learning-by-doing.

This paper takes the existence of such learning-by-doing as a starting point, and investigates its implications for the dynamics of innovation. I adopt an entrepreneurial slant, taking it as given that many good ideas for new, higher-quality products are naturally generated outside the established firm sector.[3] I then ask: if established firms have a countervailing advantage over newcomers in the form of an existing customer base, how is innovative activity by these newcomers affected?

More precisely, I construct a dynamic model with the following features: (1) incumbent firms' accumulated learning-by-doing gives them a firm- and product-specific "distribution cost" advantage over potential entrants, an advantage which gets more pronounced the longer the incumbents are able to survive; (2) potential entrants on the other hand, may have access to a technology for a superior new product; and (3) technological innovations have spillovers—once an entrant develops a new technology and introduces its product to the marketplace, future generations of innovators can learn from it, and improve

in nature. Moreover, large established firms may have certain organizational advantages in doing incremental innovation, as stressed by Schumpeter (1942)—e.g., preferential access to information or trained scientific personnel. Combining these strategic and organizational arguments, Henderson (1993) concludes that: "established firms are likely to dominate incremental innovation, while entrants are likely to dominate radical innovation" (p. 252). Thus if most innovation proceeds at an incremental pace, one might expect that established firms would naturally be able to stay on the cutting edge.

2. This quote is taken from "Continental Bank Roundtable on Global Competition in the '90's", *Journal of Applied Corporate Finance*, Spring 1993 (p. 40).

3. As the references in footnote 1 suggest, this entrepreneurial premise may be most relevant when the ideas in question represent radical, rather than incremental innovations. The premise is endogenized to some extent in Section 3.2 below, where I examine the different incentives that entrants and established firms have to engage in the sort of research that generates ideas for new products.

upon it.[4] The model has implications for how customer bases affect both the long-run average level of innovative activity, as well as for the temporal pattern of innovation.

In its focus on the long-run average level of innovation, the model is similar in spirit to the recent work in the growth literature that also analyses repeated innovations with intertemporal knowledge spillovers.[5] A common theme of this work is that, due to the spillovers, the resources devoted to innovation, and hence the rate of growth of the economy, can in some circumstances be less than socially optimal. That basic under-innovation effect is at work in this model too, although it can be dramatically amplified by the presence of customer bases. Loosely speaking, in my model firms compete on two dimensions: a "high-spillover" dimension—i.e., product quality—where one firm's innovation is ultimately passed on to other firms; and a "zero-spillover" dimension—distribution costs—where an incumbent's advantage never spills over to other firms. In some circumstances, an incumbent can survive a very long time (or even forever) by being strong in the zero-spillover dimension. While this is privately optimal for the incumbent, it can be very costly for society, as generation after generation of potentially high-spillover innovators are warded off.

Where the model departs more sharply from the papers mentioned above is in its emphasis on the *timing* of innovation. Most of these other works have the feature that innovation proceeds at a steady rate.[6] To take just one example, in Aghion and Howitt (1992)—which apart from the customer bases resembles this model quite closely—there always exists a steady-state equilibrium in which innovative activity is constant over time. In contrast, a central feature of the model in this paper is that when a new firm successfully displaces an incumbent in any given period, this has a positive externality on future generations of potential entrants—it makes it easier for them to gain access to the market. This externality, which I call the "shakeup" externality, arises because when a new firm succeeds, it breaks the incumbent's stranglehold on the customer base. Thus the market is now "up for grabs", which tilts the playing field more toward technologically strong newcomers, and away from established firms.

The consequence of this shakeup externality is that even if the underlying research and development technology is stable over time, innovations will tend to occur in waves. That is, if there is an innovation today, the odds of another innovation tomorrow may be substantially higher. Thus on the one hand, there may be long periods of stagnation, in which no newcomers enter the market. But these periods of stagnation can give way to rapid bursts of innovative activity. This wave-like aspect of innovation was also stressed by Schumpeter (1936):

> "...new combinations are not, as one would expect according to general principles of probability, evenly distributed through time ... but appear, if at all, discontinuously in groups or swarms .... Why do entrepreneurs appear, not continuously, that is

4. This description of the model makes it sound superficially like that of Young (1993), who also studies the interaction of invention and bounded product-specific learning-by-doing. However, in Young (1993), as in much of the growth literature, the learning-by-doing is *not* firm-specific; rather, it spills over across sectors. Thus learning-by-doing does not create entry barriers, and the sorts of issues I am interested in do not arise.

5. In addition to the papers cited above, see also Romer (1990).

6. See, however, Shleifer (1986) for a model where innovations are implemented in bunches. The mechanism in Shleifer's model is quite different from that emphasized here—bunching is driven by aggregate demand spillovers that lead firms to coordinate their implementation in periods when demand is high. In contrast, demand plays no role in my model; everything is driven by the relative competitive strengths of the rival firms. Also somewhat related is Jovanovic and MacDonald (1994*a*). Although there is no endogenous bunching of *innovation* in their model, large innovations increase the incentives for laggard firms to engage in costly *imitation*. This gives rise to what they term "waves of change and improvement" (p. 26).

singly in every appropriately chosen interval, but in clusters? Exclusively because the appearance of one or a few entrepreneurs facilitates the appearance of other, and these the appearance of more, in ever-increasing numbers." (pp. 223, 228)

As this passage indicates, Schumpeter also believed that the entry of any one entrepreneur imparted a positive externality to future would-be entrepreneurs, and that this was the source of waves, or swarms, in innovative activity. However, he was somewhat less clear as to exactly what the sources of this externality were. One major contribution of this paper lies in delineating more precisely the mechanisms that give rise to this positive externality.

The remainder of the paper is organized as follows. The basic model is developed in Section 2. For the purposes of this section, it is just assumed that inventions arrive exogenously to potential entrants, who must then decide whether or not to spend the money to further develop these inventions and enter into product market competition against incumbent firms. In Section 3, I extend the model to allow both entrants and incumbents to choose an optimal level of research activity. This then endogenizes the probability that the entrant will have sole access to a new invention. As will be seen, this extension of the model in many cases further strengthens the positive "shakeup" externality effect, and leads to an even more pronounced bunching of innovative activity. Section 4 concludes.

2. THE MODEL

2.1. *Types of firms*

The model is an infinite horizon one. In each period t, there is a single "incumbent" firm, defined as the firm that was active in period $t-1$. With probability p, there is also a "potential entrant" firm, that has sole access to the technology for a new product, and that may choose to further develop the technology and challenge the incumbent for the market in period t. For the time being, p is taken as exogenous; later it will be made an endogenous function of the resources devoted to research by both the entrant and the incumbent.

If the potential entrant is successful in its challenge, it will become the new incumbent in period $t+1$. Only one firm is ever active in any given period—an incumbent always has 100% market share. As will be seen, this feature emerges from the assumptions that are made below about firms' cost structures and the nature of product market competition.

In addition to incumbents and potential entrants, there is a third class of firms, called "copycats". As will be seen below, copycats are always strictly less efficient than incumbents. Thus in equilibrium, they never capture any market share. Nonetheless, they play an important role: their costs serve to tie down market prices in periods when the incumbent firm is not challenged by a potential entrant, and otherwise would be an uncontested monopolist.

2.1. A *Incumbents' cost structures*

The incumbent, the potential entrant and the copycats all have different cost structures; I begin by describing that of the incumbent. The incumbent has no fixed costs. Its marginal costs are of two kinds: costs of "production" and costs of "distribution". The marginal cost of production for an incumbent at time t is denoted by C_t^i. For any given incumbent, production costs do not change over time. That is, if the same firm is the incumbent in periods $t-1$ and t, and it produces the same product, then $C_t^i = C_{t-1}^i$.

The marginal cost of distribution for an incumbent at time t is denoted by D_t^i. The costs of distribution are assumed to take the following form:

$$D_t^i = dC_t^i \beta^{A_t^i} \quad \text{for } A_t^i \leq \bar{A};$$
$$= dC_t^i \beta^{\bar{A}} \quad \text{for } A_t^i > \bar{A} \tag{1}$$

where $\beta < 1$ and A_t^i is the "age" of the incumbent at time t—that is, A_t^i represents the number of consecutive periods (prior to t) over which the incumbent has been actively selling the same product.

There are two things to note about the form of the incumbent's distribution costs in (1). The first is that they are proportional to the incumbent's production costs. This assumption is not really critical for the basic point to be made. However, it makes it possible to couch the analysis in a steady-state equilibrium framework. Indeed, in order to facilitate the steady-state approach, all costs incurred by any firm in the model will be proportional to that firm's contemporaneous production costs. As will become clear shortly, this simplification allows one to derive a set of time-invariant decision rules for firms—rules that do not depend on, for example, a firm's current ratio of production costs to distribution costs.

The second, more important assumption embedded in (1) is that an incumbent firm's distribution costs are a decreasing function of the length of time it has been selling the same product. In particular, distribution costs decline geometrically over the first $\bar{A}$ periods of incumbency and then remain flat after that. While the specific functional form is not critical, it is crucial that these costs do decline over some range. It is not enough simply to assume—as would be implied by a switching cost model—that an incumbent has lower distribution costs than an entrant, but that all incumbents have the same costs regardless of age. It must be the case that *the longer* an incumbent has been around, the stronger their competitive advantage along this dimension.

How should this assumption— that distribution costs decline with an incumbent firm's age—be interpreted? One possibility is to think of falling distribution costs as reflecting the accumulated knowledge about customers that is a by-product of an ongoing firm-customer relationship.[7] For example, the better a computer manufacturer knows its corporate customers, the more efficiently it can market to them, customize the computers it sells them, provide continuing service that meets their specific needs, etc.

As stressed in the Introduction, what is important about the distribution costs is that they represent a competitive attribute where there is effectively the equivalent of learning-by-doing at the firm level, but *no spillovers* across firms. As will become clear, it is this lack of spillovers across firms that differentiates distribution costs from production costs in the model. When a new production technology is pioneered by one firm, other firms are eventually able to learn from it and improve upon it. In contrast, firms are unable to inherit the distribution cost advantages created by their predecessors—if a new firm enters the market, it must effectively start from scratch in building an efficient distribution system.

In addition to being firm-specific, the learning-by-doing is also product-specific. This is reflected in the assumption that A_t^i is a measure of the number of periods the incumbent has been producing a single particular product. If the incumbent tries to switch products—i.e., if it tries to knock off an innovation coming from an entrant—its A_t^i reverts back to

7. In this sense, the value of long-term relationships is similar to that which has been emphasized in the literature on banking. Bank relationships are valuable, this literature argues, because of the accumulated knowledge about borrowers that they generate. See, e.g., Fama (1985), Sharpe (1990), Rajan (1992), and Petersen and Rajan (1994), among others.

zero. Continuing with the computer example, if an established mainframe manufacturer suddenly starts trying to sell PC's to its customers, its experience in selling mainframes does not give it an advantage over a new entrant to the PC market.[8]

2.1.B. *Potential entrants' cost structures*

It is assumed that the potential entrant has invented a new technology, which if "developed", will allow it to produce at a lower quality-adjusted cost than the incumbent. Henceforth, I will be a bit loose and refer to the new product as simply having lower "production costs" than the existing one. However, it is important for the logic of the model that one thinks of of the new product as being distinct from the existing one—although it serves the same function and hence is a substitute for the existing one from the perspective of consumers, it is not literally the same item made more cheaply.[9] Again, the computer example is helpful here: one can think of the innovation of PC's as allowing certain kinds of computing services to be delivered more efficiently than with mainframes, but PC's and mainframes are distinct products. This is the "quality ladder" interpretation of innovation adopted by Aghion and Howitt (1992), Grossman and Helpman (1991*a*, *b*) and others.

The potential entrant's post-development production costs in period t, C_t^e, are given by:

$$C_t^e = \lambda_t C_t^i \tag{2}$$

where $\lambda_t < 1$ represents the magnitude of the innovation. λ_t is a random variable distributed according to the time-invariant cumulative density function $G(\lambda)$—the smaller is the realization of λ_t, the more significant is the period-t innovation.

A critical feature of the production cost innovations is that they exhibit spillovers across firms. In particular, if the potential entrant does indeed develop its innovation—i.e., it decides to go ahead with production in period t and takes over the market—others can learn about the new technology, beginning in period $t+1$. This will allow future generations of innovators to stand on the shoulders of the entrant, so that their innovations will further reduce production costs from a new base level of C_t^e.

Note that this assumption is already built into the notation: if a potential entrant decides to go ahead with production in period t, it will become the incumbent at time $t+1$. Therefore $C_{t+1}^i = C_t^e$ if there is entry in period t. If another innovator enters in period $t+1$, that second innovator will improve on the costs of the period $t+1$ incumbent, so that $C_{t+1}^e = \lambda_{t+1} C_{t+1}^i = \lambda_{t+1}\lambda_t C_t^i$. Thus if there are two consecutive rounds of innovation followed by development, production costs will fall by a factor of $\lambda_{t+1}\lambda_t$. More generally, if there are n rounds of innovation and development in any given interval, production costs will fall by a factor equal to the product of all the λ's that were developed.

While the potential entrant has an edge over the incumbent in terms of production costs, it is at a disadvantage in terms of distribution costs. The potential entrant's distribution costs, denoted by D_t^e, satisfy:

$$D_t^e = dC_t^e \tag{3}$$

8. As will become clear, this also prevents incumbent firms from simply acquiring firms with innovative new products so as to sell these new products through their existing distribution channels.

9. Again, this is because I want to be able to maintain the assumption of product-specific learning-by-doing.

The form of (3) is similar to that of (1)—distribution costs are proportional to production costs for both incumbents and entrants—but the ratio of distribution costs to production costs is higher for entrants. Essentially, an entrant has the same distribution costs that would be associated with an "age zero" incumbent with the same production technology. Clearly, if $\lambda_t = 1$, and the new entrant does not have any production-cost advantage, it will be strictly less efficient overall than the incumbent.

In addition to the marginal production and distribution costs, an entrant must pay a one-time fixed development cost to begin production with the new technology. This development cost can be thought of as the amount that must be spent to turn the research discovery into a commerically viable technology. It is denoted K_t^e, and satisfies:

$$K_t^e = kC_t^e \tag{4}$$

Like the distribution costs, the development cost is proportional to the firm's current production costs. As noted above, this is done to allow for a steady-state analysis of the model.

2.1.C. *Copycats' cost structures*

In any period t, there is also a competitive fringe of copycat firms that can mimic the production technology of the current incumbent. That is, a copycat's cost of production, denoted by C_t^c, is given by:

$$C_t^c = C_t^i \tag{5}$$

However, since the copycats have no existing customer base, their distribution cost, D_t^c, is given by:

$$D_t^c = dC_t^c \tag{6}$$

Since they use only existing technology, copycat firms do not have to pay any fixed development costs. Simply put, a copycat is like an incumbent, but without the distribution cost advantage.[10] Thus its overall marginal costs are always strictly higher than those of the current incumbent.

2.2. *Product market competition*

I now turn to the competition in the product market that determines both: (1) who will control the market in any period t: and (2) the associated profits. To keep things simple, the demand side of the market is modelled in such a way as to be essentially irrelevant. This is accomplished by assuming that there are N consumers, each of whom wishes to purchase exactly $1/N$ units of the good, up to a reservation price of U_t. U_t is sufficiently large that it always exceeds the copycats' marginal costs, $C_t^c + D_t^c$. This implies that prices and profits in the model will be determined solely by the relative cost structures of the competing firms.

There are two cases to distinguish. In the first case, which occurs with probability $(1-p)$, there is no new invention in period t, and hence no potential entrant. In this case,

10. The underlying assumption here is that the production technology does not have patent protection, and hence can be costlessly imitated after a one-period lag. Although this assumption simplifies the analysis somewhat, it is of no significant consequence. Similar results would follow if one were to grant incumbents patent protection and use another device to tie down prices in those periods in which an incumbent is not challenged by a new entrant.

the incumbent competes against just the fringe of copycat firms. This implies that the incumbent, by virtue of its lower marginal cost, captures the entire market, at a price equal to the copycats' marginal cost, $C_t^c + D_t^c$. Thus the incumbent's profits in this case are given by $(C_t^c + D_t^c) - (C_t^i + D_t^i) = dC_t^i(1 - \beta^{\hat{A}_t^i})$; where $\hat{A}_t^i$ is defined as min $(A_t^i, \bar{A})$.

In the second case, which occurs with probability p, there is a potential entrant in period t. This case is a bit more complicated, and the timing of events is as follows. First, the potential entrant must decide whether or not to sink the development cost K_t^e. If it does not, then the incumbent is left alone with the copycats, and the outcome is exactly the same as described just above. If, on the other hand, the entrant does sink the development cost, then the entrant and the incumbent compete à la Bertrand. the copycats will be irrelevant to the outcome of this Bertrand competition, as their costs are now strictly higher than those of the other two types of firms.

To simplify the nature of the Bertrand competition between the incumbent and the potential entrant, I make the following assumption:

$$C_t^e + D_t^e < C_t^i + D_t^i \text{ for any } A_t^i \text{ and all } t\text{; or equivalently,}$$
$$\lambda_t < (1 + d\beta^{\bar{A}})/(1 + d) \text{ for all } t. \tag{7}$$

In words, (7) says that once development costs are sunk, the entrant's innovation is sufficiently valuable that it always has lower overall marginal costs than an incumbent *of any age*. The value of this assumption is established in the following lemma.

Lemma 1. *If (7) holds, then conditional on development, the Bertrand equilibrium always involves the potential entrant immediately gaining 100% market share in period t and charging a price in that period equal to the combined production and distribution costs of the incumbent firm,* $C_t^i + D_t^i$.

Proof. See the appendix. ||

Without the assumption that (7) holds, things might be much more complicated, although I do not think any of the basic conclusions below would be altered. To see why, suppose that $\bar{A} = 5$, and there is a three-period-old incumbent who is currently at a cost disadvantage relative to the entrant—i.e., $\lambda_t < (1 + d\beta^3)/(1 - d)$. However, suppose also that (7) is violated, such that $\lambda_t > (1 + d\beta^4)/(1 + d)$. In this situation, the incumbent might be willing to price below marginal cost today, in an effort to hang on to its customer base. While next period the potential entrant will still be around—it has already sunk the development cost—the incumbent will be in a stronger relative position, because its distribution costs will have fallen. Indeed, if the incumbent can hang on until the next period, it will have lower overall costs than the entrant and may hope to earn a profit.

The upshot is that without (7), one cannot rule out the possibility that the incumbent might engage in a complex intertemporal pricing strategy, trading off current profits in an effort to maintain its customer base. This in turn would imply that even if the entrant does take over the market in equilibrium, the price at which it does so would be lower than $C_t^i + D_t^i$, and would in general depend on the nature of the incumbent's intertemporal problem.

In contrast, if (7) holds, these issues do not arise. No matter how long the incumbent holds off the potential entrant, it will never have the lower costs. Thus there is no gain to the incumbent from pricing below current cost, and it will never do so. As a result, (7) ensures that, once the development cost K_t^e is sunk, the entrant will always take over the market completely. However, it does not guarantee that the entrant will find it worthwhile to spend the K_t^e in the first place. For that to happen, it must be that expected profits *net* of the development cost—denoted by π_t^e—exceed zero.

These profits can be broken into two components. First, there is the net profit (after development cost) earned by the entrant in the period it enters—i.e., in period t. This can be calculated directly from the results of the lemma as: $C_t^i + D_t^i - C_t^e - D_t^e - K_t^e = C_t^e[((1+d\beta^{\hat{A}_t^i})/\lambda_t) - 1 - d - k]$

In addition to these immediate profits, a firm that enters in period t can also earn profits in later periods if there is no subsequent entry by later innovators, because then it only has to contend with the weaker copycat firms.[11] Of course, if subsequent entry by an innovating firm does occur in some future period $t+j$, the period-t entrant earns nothing from $t+j$ onward.

We have already calculated the profits that a period-t entrant will earn in period $t+1$, conditional on no entry by a new innovator at this time. This profit is simply that earned by a 1-period old firm competing only with copycats, or $dC_t^e(1-\beta)$. Similarly, the profits that a period-t entrant will earn in period $t+2$, conditional on no entry by an innovator by this time, are given by $dC_t^e(1-\beta^2)$.

Thus overall, the expected present value of net profits to a potential entrant in period t is given by:

$$\pi_t^e = C_t^e\{[((1+d\beta^{\hat{A}_t^i})/\lambda_t) - 1 - d - k] + F^e\} \tag{8}$$

where F^e is a measure of the expected future profits. F^e in turn can be written as:

$$F^e = d\left\{\sum_{i=1}^{\bar{A}} \frac{(1-\beta^i)}{(1+r)^i} X_i^e + \sum_{i=\bar{A}+1}^{\infty} \frac{(1-\beta^{\bar{A}})}{(1+r)^i} X_i^e\right\} \tag{9}$$

where r is the per-period interest rate and where X_i^e is defined as the probability that the entrant will not be displaced within i periods by a future innovator. (I will show how X_i^e is explicitly calculated momentarily.)

2.3. *Steady-state equilibrium development rules*

The appropriate equilibrium concept for this model is a set of equilibrium development rules. In particular, for any given age of the incumbent firm A, we want to calculate a "threshold" $\lambda^*(A)$ such that a potential entrant will choose to develop its invention if and only if $\lambda_t < \lambda^*(\hat{A}_t^i)$. Intuitively, in this equilibrium, a potential entrant will weigh both the magnitude of its own innovation, and the age of the current incumbent, in making the development decision. At the same time, it will take as given the set of equilibrium development rules when it attempts to assess the likelihood of entry by future generations of innovators—i.e., when it calculates the values of X_i^e.

For example, if the potential entrant does decide to go forward in period t, the probability that there will be another round of innovation *and* entry in period $t+1$ is given by: $pG(\lambda^*(1))$. Similarly, the probability of another round of innovation and entry in period $t+2$ (conditional on there being no entry in period $t+1$) is $pG(\lambda^*(2))$. Thus we

11. Note that once an entrant has eliminated an incumbent in period t, it does not have to worry about this ex-incumbent imitating the new technology and using its existing customer base to launch a comeback in some period $t+j$. This is because of the assumption that the learning-by-doing underlying the customer base is product-specific. Once a new technology has been adopted, an ex-incumbent is no better off than a copycat firm that has never acquired any learning-by-doing experience.

have the following expressions for X_i^e:

$$\begin{aligned} X_i^e &= \prod_{j=1}^{i} (1 - pG(\lambda^*(j))) \quad \text{for } i \leqq \bar{A} \\ &= \{\prod_{j=1}^{\bar{A}} (1 - pG(\lambda^*(j)))\}(1 - pG(\lambda^*(\bar{A})))^{(i-\bar{A})} \quad \text{for } i > \bar{A} \end{aligned} \tag{10}$$

The important observation that follows from (10) is that F^e depends on the equilibrium rules $\lambda^*(1), \ldots, \lambda^*(\bar{A})$, but is independent of the period-t data, λ_t and $\hat{A}_t^i$. Moreover, it is easy to show that F^e is a monotonically decreasing function of each of the λ^*'s. This makes intuitive sense: the greater is any λ^*, the greater is the probability that a future innovator will decide to enter, and hence the lower are the expected profits associated with entry today. This is similar to the "creative destruction" effect noted by Aghion and Howitt (1992) and others.

The λ^*'s are determined by setting the appropriate variants of (8) equal to zero. For example, $\lambda^*(1)$ satisfies:

$$[((1 + d\beta)/\lambda^*(1)) - 1 - d - k] + F^e = 0 \tag{11}$$

Analogously, $\lambda^*(2)$ satisfies:

$$[((1 + d\beta^2)/\lambda^*(2)) - 1 - d - k] + F^e = 0 \tag{12}$$

Given that F^e is independent of λ_t and $\hat{A}_t^i$, we can use equations like (11) and (12) to derive the following recursive relationship among the λ^*'s:

$$\lambda^*(A)/\lambda^*(1) = (1 + d\beta^A)/(1 + d\beta); \quad \text{for all } A \leqq \bar{A} \tag{13}$$

In order to solve explicitly for the individual λ^*'s, we need to pin down a "boundary condition"—i.e., we need to establish the value of, say $\lambda^*(1)$. While this is difficult to do in closed form, it is easy to establish the existence and uniqueness of a solution, and to do some simple comparative statistics. From (11), $\lambda^*(1)$ is determined by the following equality:

$$1 + d + k - (1 + d\beta)/\lambda^*(1) = F^e \tag{11'}$$

We have already seen that the term on the right-hand side of (11'), F^e is a decreasing function of $\lambda^*(1)$. It is also easy to see that the term on the left-hand side of (11') is an increasing function of $\lambda^*(1)$. Thus the unique equilibrium value of $\lambda^*(1)$ will be determined as in Figure 1. In the figure, the increasing function $1 + d + k - (1 + d\beta)/\lambda^*(1)$ is denoted by "Current", since this function represents the net current (i.e., period t) *losses* associated with entry when $\lambda_t = \lambda^*(1)$. The decreasing function F^e is denoted by "Future". Thus the equilibrium point has the simple interpretation of being that value of λ_t where the current costs of entry are just equal to the expected future profits.

Example 1. In this example, the parameter values are chosen as follows: $p = 1$; $d = 1$; $k = 5$; $\beta = 0{\cdot}8$; $\bar{A} = 10$; and $r = 0{\cdot}10$. In addition, the distribution $G(\lambda)$ is assumed to be uniform over the interval (0, 0·5). It is easily checked that this distribution satisfies the condition in (7).

These parameters lead to a value of $\lambda^*(1) = 0{\cdot}268$.[12] From (13), if follows that the λ^*'s then decline monotonically to a value of 0·165 for $\lambda^*(10)$. The probability of a

12. As in all the examples that follow, $\lambda^*(1)$ is computed using an iterative algorithm. I begin with a trial value. This trial value can be used to calculate all the other λ^*'s from (13), and then F^e from (9) and (10). I then substitute the resulting F^e into equation (11), and generate a second-round value of $\lambda^*(1)$. The entire procedure is then repeated until convergence.

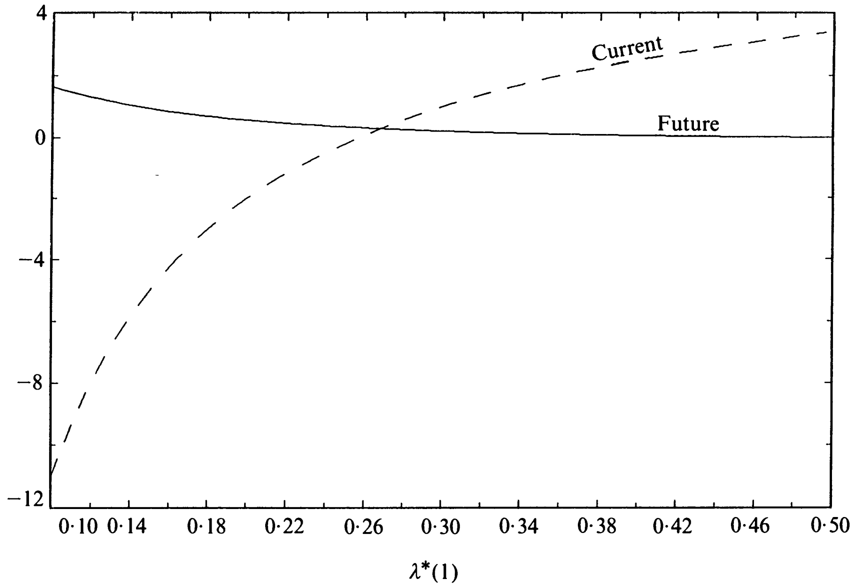

FIGURE 1

Costs and benefits of entering market

developed innovation when the incumbent is only one period old is $0.268/0{\cdot}5 = 0{\cdot}536$. However, when the incumbent is ten or more periods old, this probability falls to $0{\cdot}165/0{\cdot}5 = 0{\cdot}330$. It is also possible to calculate the unconditional, steady-state probability of a developed innovation, which in this example is equal to 0·491.

2.4. *Implications for the pace of innovation*

One result that emerges immediately from Figure 1 has to do with the effect of development costs on the pace of innovation:

Proposition 1. *An increase in the ratio k of development costs to production costs reduces the threshold λ^*'s, thereby deterring the development of new innovations.*

Proof. This follows from differentiating (11), and observing that $dF^e/d\lambda^*(1) < 0$. ||

Although the impact of increased development costs on innovation is unambiguous in sign, its magnitude is dampened somewhat because there are two competing effects at work. On the one hand, an increase in k has a direct negative impact on the profits earned by an entrant in the period these costs are paid—i.e., in the period that entry occurs. This is manifested as an inward shift of the upwards-sloping Current curve in the figure. On the other hand, an increase in development costs *raises* the expected profits that an entrant will earn in the periods *after* entry, because it lowers the probability of further rounds of entry by subsequent innovators. This is reflected in the fact that the Current curve shifts *along* the downwards-sloping Future curve, rather than along a horizontal line.

Example 2. Maintain all the same parameter values as in Example 1, but double k, from 5 to 10. $\lambda^*(1)$ falls from its previous value of 0·268 to 0·161. Thus the probability of a developed innovation when the incumbent is one period old falls from 0·536 to 0·322. The unconditional probability of a developed innovation falls from 0·491 to 0·269.

More central to this paper are the effects of customer bases on innovative activity. A first natural question to ask is: as customer bases become relatively more important in inter-firm competition, does the pace of innovation tend to accelerate or slow down? That is, does the unconditional probability of a developed innovation rise or fall?

There are two ways to parametrize the importance of customer bases: either an increase in d or a decrease in β can be thought of as raising the competitive advantage of incumbent firms along this dimension. (In the extreme cases where either $d=0$ or $\beta=1$, incumbents never have any advantage.) Whichever measure is used, the following result obtains:

Proposition 2. *The effect of customer bases on the unconditional probability of innovation is in general ambiguous. In some cases, an increase in the importance of customer bases (i.e., an increase in d, or a reduction in β) can dramatically stifle innovation. However, there are other cases in which an increase in the importance of customer bases can actually promote the development of innovative new technologies.*

The proposition can be established with a couple of examples. Before turning to the examples, it is useful to demonstrate heuristically the sources of the ambiguity. This can be done by reference to Figure 1. For concreteness, consider the impact of a decrease in β. On the one hand, this reduces the current-period appeal of entry, thereby leading to an inward shift of the Current curve. At the same time, once entry has occurred, a lower value of β makes future profits higher in each period in which the new entrant remains in control of the market and faces only the copycat firms. This causes an outward shift of the Future curve. The net effect of these two shifts on $\lambda^*(1)$ cannot in general be signed. However, the latter effect is more significant when p is low, because then new inventions arise infrequently and it is more likely that the higher post-entry profits can be sustained for a longer period of time. This suggests that customer bases are relatively more favourable to innovation when p is close to zero.

Example 3. The potential for customer bases to have a negative impact on innovation can be illustrated most starkly by considering a case where: (1) $(1+d)/(1+d+k)>\lambda^*(\bar{A})$ (this can always be accomplished by making d and $\bar{A}$ large enough); (2) the distribution $G(\lambda)$is such that λ_t always satisfies $\lambda^*(\bar{A})<\lambda_t<(1+d)/(1+d+k)$; and (3) p is strictly less than one. In this case, the economy eventually must get "stuck" in a situation where the incumbent firm is at least $\bar{A}$ periods old, and where there are never any more new innovations developed. All it takes to get an incumbent to be this old is a run of $\bar{A}$ consecutive periods with no new inventions; with $p<1$, such a run will occur in finite time with probability one. And once the incumbent reaches this age, there is no innovation that is of sufficient magnitude to displace it, since it is always the case that $\lambda^*(\bar{A})<\lambda_t$.

In contrast, if we maintained the same $G(\lambda)$, but set $\beta=1$, so that customer bases did not matter, the economy could never get stuck in this way. Rather, *every* innovation that arose would always be developed. This is guaranteed by the assumption that $\lambda_t<(1+d)/(1+d+k)$, which makes entry immediately profitable against an incumbent with no distribution cost advantage. The net result is a new developed innovation in a fraction p of the periods.

Example 4. To see how customer bases can actually *help* foster innovation when p is low, consider a case where the parameter values are all the same as in Example 1, except that $p=0{\cdot}10$ instead of 1. In this case, the unconditional probability of a developed innovation is 0·064. If, however, we take these same parameters but set $\beta=1$ (keeping $p=0{\cdot}10$) the unconditional probability falls somewhat, to 0·057. Thus there is actually more innovation when customer bases matter than when they do not.

Although the examples are useful in highlighting the effects that go in either direction, it would be nice if one could derive explicit necessary and sufficient conditions for, e.g., an increase in customer bases to have a positive effect on the probability of innovation. Unfortunately, this is difficult to do in general—deriving the necessary and sufficient conditions as a function only of exogenous parameters is effectively tantamount to solving the model in closed form. However, in one important limiting case, it is possible to make a simple and intuitive statement:

Proposition 3. *In the limiting case where p approaches zero, a sufficient condition for an increase in customer bases (i.e., an increase in d, or a reduction in β) to increase the unconditional probability of innovation is*:

$$\sum_{i=1}^{\bar{A}} \frac{(1-\beta^i)}{(1+r)^i} + \sum_{i=\bar{A}+1}^{\infty} \frac{(1-\beta^{\bar{A}})}{(1+r)^i} > (1-\beta^{\bar{A}}(1+k)). \tag{14}$$

Proof. See the appendix. ||

Since the sufficient condition in (14) is very likely to be satisfied—indeed, the only way that it can be violated is by making the interest rate r extremely high—Proposition 3 confirms the basic intuition suggested by Example 4; namely that customer bases are likely to be favourable to innovation when p is low. It is worth emphasizing that this outcome stands in contrast to most static models of entry prevention in the industrial organization literature. (see Tirole (1988, Chapter 8) for a detailed overview and references.) In standard one-shot models, the ability of incumbents to accumulate specific learning-by-doing serves to unambiguously deter entry. What is different here is the dynamic nature of the model: today's potential entrant recognizes that if it does indeed enter, it will become tomorrow's incumbent, in which case it will actually derive future benefits from the existence of learning-by-doing entry barriers.

To this point, the examples and propositions have focused on the effect that customer bases have on the long-run average level of innovation. But perhaps the most novel aspect of the model has to do with its implications for the *timing* of innovations:

Proposition 4. *Customer bases lead to an endogenous "bunching" of innovative activity. In particular, whenever there is an invention of sufficient merit to be developed in period t, this raises the likelihood of further development in future periods.*

This result follows immediately from the recursive formula for the λ^*'s in equation (13), which states that, the older is the current incumbent firm, the less attractive it is for a potential entrant to develop a given invention. This observation in turn implies that when a potential entrant does go forward in period t, this has a *positive externality on future generations of potential entrants*—it raises their proposective returns to development, since they will be facing a younger incumbent, on average. This positive externality might be termed a "shakeup" externality, since its essence is that entry in period t breaks an existing incumbent's hold on customers, thereby shaking up the market and facilitating the entry of young firms whose competitive advantage lies more in technological prowess and less in having an established customer base.

Example 5. Maintain all the same parameter values as in Example 1. Suppose there is an incumbent firm in period t that is over $\bar{A}$ periods old, and that a potential entrant is deciding whether or not to proceed with development. Let us ask what effect the potential entrant's decision has on the probability of development in future periods, $t+1$, $t+2$, etc.

If the potential entrant does not develop its invention in period t, the new potential entrant in period $t+1$ will still be facing an incumbent over $\bar{A}$ period old. Thus the probability of a developed innovation in period $t+1$ is $pG(\lambda^*(\bar{A}))=0{\cdot}330$. In contrast, if the potential entrant does proceed with development in period t, the new potential entrant in period $t+1$ will have an easier task, since it will be facing an incumbent only 1 period old. Thus the probability of development in period $t+1$ rises to $pG(\lambda^*(1))=0{\cdot}536$.

The potential entrant's decision in period t exerts a similar, though less pronounced influence on the conditional probabilities for period $t+2$. If there is no development in period t, the probability of development in period $t+2$ is given by: $(1-pG(\lambda^*(\bar{A})))(pG(\lambda^*(\bar{A})))+(pG(\lambda^*(\bar{A})))(pG(\lambda^*(1)))=0{\cdot}398$. If, on the other hand, there is development in period t, the probability of development in period $t+2$ rises to: $(1-pG(\lambda^*(1)))(pG(\lambda^*(2)))+(pG(\lambda^*(1)))^2=0{\cdot}514$.

More generally, given that there is development in period t, the conditional probabilities of development in further-out periods decay slowly back to the unconditional value of 0·491. For example, the conditional probability at $t+3$ is 0·504; at $t+4$ it is 0·499, etc. Figure 2 illustrates this time path of conditional probabilities.

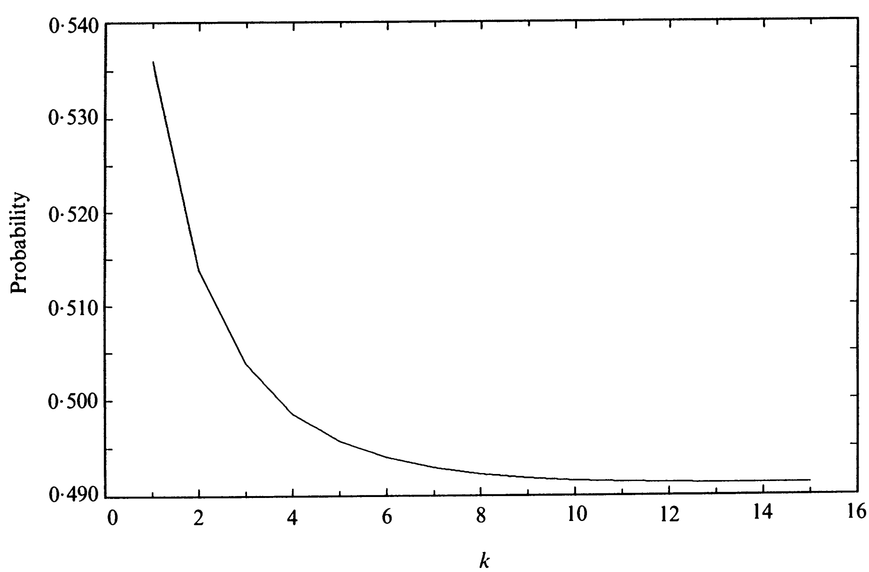

FIGURE 2

Probability of innovation at time $t+k$
Conditional on innovation at time t

3. ENDOGENOUS RESEARCH EXPENDITURES

Thus far, the probability p of an invention arriving in any given period has been taken as exogenous. Additionally, it has been assumed that only potential entrants ever generate inventions; incumbents never do. More realistically, the probability that either a potential entrant or an incumbent has access to a new technology will be a function of the resources that each devotes to research activities. These resources will in turn depend on their respective incentives to engage in research. In light of this motivation, I now extend the model to incorporate endogenous research expenditures by both entrants and incumbents. I do this in two steps. First, I consider the simpler case where only entrants do research. Next I add incumbent research into the mix. To preview the results a bit, it turns out that endogenous research can—though this need not be so under all circumstances—greatly amplify the tendency for innovations to come in waves. The intuition is simple: a shakeup of the market in period t now not only makes it more attractive for a potential entrant to develop an existing invention in period $t+1$, it also will encourage more research by potential entrants in period $t+1$, and hence *raise the likelihood of there being an invention in the first place.*

3.1. *Only potential entrants do research*

Assume as before, that there is just one potential entrant in each period. The potential entrant now begins the period by deciding how much effort to expend on research. The probability that the research will be successful and yield an invention in period t is then given by $p(e_t)$, where e_t is the level of research effort, and $p(\,)$ is an increasing, concave function. For some purposes, it will be useful to recognize that even with infinite effort, $p(e_t)$ is bounded strictly away from one—i.e., it asymptotes to $p^{\max}<1$. This captures the idea that no matter how hard one tries, research is an inherently uncertain proposition.

Research effort is costly to the potential entrant; in particular, the cost of an effort level e_t is given by R_t^e, which satisfies:

$$R_t^e = e_t C_t^i \tag{15}$$

As with all the other costs in the model, research costs are proportional to contemporaneous production costs. The only slight twist is that the potential entrant's research costs are proportional to the *incumbent's* current production costs. This simplifies the interpretation, since one can think of the incumbent's production costs as already known at the time that the potential entrant begins research.[13]

In this extended version of the model, an equilibrium will consist of two sets of decision rules. First, as before, there will be the threshold λ^*'s that tell us when development is optimal as a function of the incumbent's age. Second, there will be a set of optimal research levels, denoted by $e^*(1), \ldots, e^*(\bar{A})$, which tell us how much research will be done as a function of the incumbent's age.

In order to begin thinking about the properties of this equilibrium, let us first focus on equilibrium at the development stage. That is, take the research rules—the e^*'s—as given, and solve for the λ^*'s. To do so, note that equations (8)–(10), which describes an entrant's profits from development, are still valid. The only difference is that X_i^e, and hence F^e, are now more complicated functions since they depend also on the e^*'s. It

13. In contrast, it is a little more awkward to assume that the entrant's research costs are proportional to its *own* production costs, since these will not be known until the research is completed. Nonetheless, using this alternative assumption would have no effect on the results below.

follows immediately that the recursive equation (13) for the λ^*'s is exactly the same as before. In other words, once we pin down $\lambda^*(1)$, the ratios of the λ^*'s will be the same as they were earlier.

The computation of $\lambda^*(1)$ itself is slightly more complicated, but the logic seen earlier still applies exactly. That is, for any given set of e^*'s, one can still show that there is a unique equilibrium value of $\lambda^*(1)$ as in Figure 1. The only change is that the Future curve in this figure, while still downwards sloping, may have a slightly different shape, due to its more complicated functional form.

Having solved for the equilibrium development rules, one can now fold backwards and solve for the equilibrium research levels. Expected profits at the research stage, denoted by π_t^r, satisfy:

$$\pi_t^r = C_t^i \left\{ p(e_t) \left[\int_{\lambda < \lambda^*(\hat{A}_t^i)} \frac{\pi_t^e(\hat{A}_t^i, \lambda_t) g(\lambda) d\lambda}{C_t^i} \right] - e_t \right\}. \tag{16}$$

Here $\pi_t^e(\hat{A}_t^i, \lambda_t)$ is the profit at the *development* stage for a given value of $\hat{A}_t^i$ and λ_t. It is calculated using equations (8)–(10).

From equation (16), one can immediately derive the first order conditions that must be satisfied by each of the e^*'s:

$$\begin{aligned} p'(e^*(A)) &= \left[\int_{\lambda < \lambda^*(A)} \frac{\pi_t^e(A, \lambda) g(\lambda) d\lambda}{C_t^i} \right]^{-1} \\ &= \left[\int_{\lambda < \lambda^*(A)} (1 + d\beta^A - \lambda(1 + d + k) + \lambda F^e) g(\lambda) d\lambda \right]^{-1} \\ &\equiv 1/E(A) \end{aligned} \tag{17}$$

Equation (17) makes it clear that, in any equilibrium, $e^*(A)$ must be a decreasing function of A.[14] This follows from the fact that both the integrand in (17), as well as the limit of integration, $\lambda^*(A)$, are decreasing functions of A. Intuitively, the older the incumbent, the lower is the marginal productivity of investment in research by a potential entrant. This is true for two related reasons. First, development of any given invention is less profitable with an older incumbent. Second, fewer inventions are worth developing at all with an older incumbent. The bottom line is that less research is done when the incumbent is older. This observation leads to the following result:

Proposition 5. *When the potential entrant's research level is made endogenous, and the incumbent continues to do no research, the bunching of innovations becomes more pronounced. That is, the positive effect of an innovation in period t on the conditional probability of innovation in period t + 1 is stronger when the potential entrant's research is endogenous.*

The reasoning for this result is straightforward. The probability of a developed innovation in period t+1 is $p(e^*(\hat{A}_{t+1}^i))G(\lambda^*(\hat{A}_{t+1}^i))$. When there is an innovation in period t,

14. Moreover, it can be shown that there exists a unique equilibrium. To see heuristically why this is so, think of drawing the following two curves on a graph whose axes are $\lambda^*(1)$ and $e^*(1)$. The first curve represents those combinations of $\lambda^*(1)$ and $e^*(1)$ that are consistent with equilibrium in the development stage. It is easy to see that this curve must be downward sloping—i.e., entry today is less attractive if future generations of entrants do more research. The second curve represents those combinations that are consistent with equilibrium in the research stage. This curve is upward sloping. This is because the only way to elicit more research today is if the appeal of entry is greater, which would imply a higher value of $\lambda^*(1)$. The intersection of the two curves pins down the unique values of $\lambda^*(1)$ and $e^*(1)$.

this has the effect of setting $\hat{A}^i_{t+1}=1$. With endogenous research, this reduction in the age of the incumbent has two distinct beneficial effects: it increases both the probability of an invention, and the likelihood that the invention will subsequently be developed. In contrast, in the simpler model of Section 2, only the latter effect was at work.

Example 6. Maintain all the same parameter values as Example 1, except: the probability p of an invention can be either 1 or 0·5, depending on the level of research effort. In any period t, it costs an amount equal to $0{\cdot}25C^i_t$ to set $p=1$; it is costless to set $p=0{\cdot}5$.

It can be shown that these parameters lead to $\lambda^*(1)=0{\cdot}28$. Moreover, when $A^i_t=1$, the expected profits conditional on obtaining an invention in period t are equal to $0{\cdot}505C^i_t$. Thus it is (just barely) worth it to spend the money on research to raise the probability of an invention from 0·5 to 1. In contrast, when A^i_t exceeds 1, it is never worth it to invest in research.

We can now revisit the question asked in Example 5, namely how does development in period t affect the conditional probability of development in period $t+1$? If we begin period t with an age-10 or older incumbent and there is no development, the conditional probability for $t+1$ is only $0{\cdot}5G(\lambda^*(10))=0{\cdot}17$. In contrast, if there is development in period t, the conditional probability for $t+1$ rises to $G(\lambda^*(1))=0{\cdot}56$.

3.2. *Both entrants and incumbents do research*

Now assume that both the potential entrant and the incumbent have access to the same research technology $p(\cdot)$, and denote their respective levels of research effort in period t by e^e_t and e^i_t. Assume further that whether or not the entrant's research is successful is independent of whether the incumbent's is, and vice versa. Thus there are three scenarios to consider in which there is at least one invention.

First, with probability $p(e^e_t)(1-p(e^i_t))$, only the potential entrant succeeds in generating an invention. This is the same case as has been analyzed throughout, and the entrant's decision of whether or not to proceed with development continues to hinge on exactly the same considerations as before.

Second, with probability $p(e^i_t)(1+p(e^e_t))$, only the incumbent succeeds in generating an invention. Conditional on having sole access to an invention, the incumbent's incentive to proceed with development is much less than the entrant's. This is a consequence of the so-called "replacement effect". (See, e.g., Tirole (1988, Chapter 10).) If the incumbent develops the invention, it cannibalizes all the profits flowing from its current product and the associated product-specific customer base.

To simplify the analysis of this case, I assume that $\lambda_t>(1+d\beta)/(1+d+k)$. This just says that no innovation is so good that is completely covers its fixed costs of development in the period in which it is introduced—a fairly mild restriction. But this weak assumption ensures that an incumbent will never proceed with development when it alone has an inventions. This is because switching to a new product always reduces the incumbent's profits in all future periods after innovation, by diminishing its advantage over the copycat firms.

Finally, with probability $p(e^i_t)p(e^e_t)$, both the entrant and the incumbent generate inventions simultaneously. In this case, I assume that they both discover literally the same invention—i.e., λ_t is the same for both. I also assume that if either one sinks the fixed development cost, the other can immediately and costlessly imitate, *without a one-period lag*. In other words, the advantage to having actually invented a new technology at the same time as another firm is that—unlike a copycat—one already has much of the know-how in place, and can thus produce the developed version of it more quickly.

These assumptions make analysis of the simultaneous-invention case very simple. Clearly, neither firm will ever proceed with development. For if one firm did spend the fixed development cost, its rival would immediately imitate it. They would then play Bertrand with equal costs, driving profits to zero. Thus the firm doing the development would never be able to recoup its fixed cost, and would always lose money.

This logic implies that a potential entrant can never benefit from an invention if it occurs at the same time as an incumbent invention. The converse is not true, however. Even though an incumbent never actually *develops* a simultaneous invention, such an invention does have value to the incumbent. To see why, note that when $\lambda_t < \lambda^*(\hat{A}_t^i)$, a potential entrant with sole access to the invention would proceed with development, enter and destroy the incumbent's profits from its existing product. By simultaneously inventing, the incumbent deters the potential entrant from going ahead with development, and thereby safeguards the profits coming from the incumbent's existing product. In other words, simultaneous invention by the incumbent can serve a valuable defensive purpose.[15]

Analogous to previous notation, we can express the profits thus protected by the incumbent in the form $C_t^i I(A_t^i)$. The function $I(\ )$ has a complicated structure, because it is recursive in nature—the value to an age-A incumbent of staying in the market in period t is that it gets the flow of period-t profits, plus the opportunity to try (as an age-$A+1$ incumbent) to optimally defend itself and stay in the market in period $t+1$, etc. In spite of this complexity, the appendix shows that the function $I(\cdot)$ is increasing with A. That is, the older is the incumbent, the larger are the expected profits from its existing product, and hence the greater is the value of safeguarding this product.

We are now ready to write down the first-order conditions for both the entrant's and the incumbent's optimal research effort decisions. The entrant's optimal effort levels satisfy:

$$p'(e^{e^*}(A)) = 1/E(A)(1-p(e^{i^*}(A))) \tag{18}$$

where $E(A)$ is a decreasing function defined as in (17) above.[16] Thus the entrant's first-order condition is basically the same as before, except that it has been adjusted by a factor $1/(1-p(e^{i^*}(A)))$. That is, to the extent that the incumbent does research in equilibrium, and hence has a non-zero probability of obtaining an invention, the entrant's incentives to do research are reduced. This makes intuitive sense, as the entrant only proceeds with development in those cases in which the incumbent does not simultaneously obtain an invention.

The incumbent's optimal effort levels satisfy:

$$p'(e^{i^*}(A)) = 1/I(A)p(e^{e^*}(A))G(\lambda^*(A)). \tag{19}$$

15. In light of this observation, I need to make one more assumption to keep the analysis manageable. I must assume that inventions are of the "use-it-or-lose-it" variety—i.e., if an invention is not developed immediately, the incumbent cannot stockpile it and retain the option to develop it at some later date. Without this assumption, the number of state variables in the problem would blow up—in order to determine an incumbent's incentive to engage in research, one would have to know not just its age, but also how many previous inventions it had already stockpiled for defensive purposes. While the assumption is admittedly unpalatable, I have no reason to believe that it changes the basic nature of the conclusions offered below.

16. The $E(A)$ function is actually a bit more complicated than in (17), because now F^e depends on the incumbent's optimal research rules. However, unlike with $I(A)$, $E(A)$ does not have a complex recursive structure. This is because if the entrant does in fact enter, the current incumbent's age A is no longer relevant in future periods. Thus in computing $E(A)$, A only matters to the extent it affects current-period profits and decisions, not future ones.

The striking difference between the incumbent and the entrant is that the incumbent does *more* research, all else equal, when the entrant does more. More precisely, the incumbent's incentive to do research is directly related to the probability that the entrant will obtain an invention worth developing, i.e., $p(e^{e^*}(A))G(\lambda^*(A))$. Again, this is because the incumbent's only motivation for undertaking research is a defensive one.

Several observations follow from these first-order conditions. First, note that as $p^{\max}$ approaches zero, the incumbent will, according to (19), put forth a vanishingly small research effort. Equation (18) says that this in turn will encourage the potential entrant to spend more heavily on research. In this limiting case, the potential entrant's research effort is determined as in the no-incumbent-research scenario of equation (17). Thus one way to rationalize the simpler formulation of most of the paper—where only the potential entrant ever generates inventions—is to say that it represents an endogenous outcome in those situations where research is inherently a long-odds bet. In such a long-odds environment, the incumbent reasons that if it ever does come up with an invention, it will almost certainly be the sole inventor. Consequently, its decision making is dominated by the replacement effect, and it optimally *chooses* to forego research activity.

Second, even if we back away from this limiting case, it is still possible to derive a couple of more general—albeit somewhat weaker—results about innovation-bunching:

Proposition 6. *With endogenous research by both the incumbent and the potential entrant, it is still the case that the probability that the potential entrant will generate an invention worth developing—given by* $p(e^{e^*}(A))G(\lambda^*(A))$*—is decreasing in the incumbent's age* A.

Proof. A complete proof is in the appendix. Here I give a heuristic version of the argument, based on the assumption that $I(A)$ is increasing in A. (This is verified in the proof in the appendix). Suppose the proposition is not true, so that $p(e^{e^*}(A))G(\lambda^*(A))$ is (weakly) increasing in A. Then, since $I(A)$ is also increasing in A, (19) tells us that the incumbent's research intensity and hence probability of success $p(e^{i^*}(A))$ must also be increasing in A. But if this is true, then since $E(A)$ is decreasing in A, (18) tells us that the potential entrant's research intensity and hence probability of success $p(e^{e^*}(A))$ must certainly be decreasing in A. And if $p(e^{e^*}(A))$ is decreasing in A, then $p(e^{e^*}(A))G(\lambda^*(A))$ must be too, since from (13) $G(\lambda^*(A))$ is decreasing in A. This establishes a contradiction and thereby proves the proposition. ||

Unfortunately, Proposition 6 is somewhat less definitive than it might at first appear, because with defensive research by the incumbent, the net probability of a new product actually making it to market is not $p(e^{e^*}(A))G(\lambda^*(A))$, but rather $(1-p(e^{i^*}(A)))$ $p(e^{e^*}(A))G(\lambda^*(A))$. It is harder to make completely general statements about how this latter quantity moves with A, since in principle, $e^{i^*}(A)$ can either increase or decrease with A.[17] However, one can make the following statement:

Proposition 7. *The net probability of a new invention coming to market—given by* $(1-p(e^{i^*}(A)))p(e^{e^*}(A))G(\lambda^*(A))$*—will always be decreasing in the incumbent's age*

17. There are two effects at work: on the one hand, as the incumbent ages, $I(A)$ goes up, so it has a greater stake in doing defensive research to protect the profits from its existing product. On the other hand, we have just seen that as the incumbent ages, the potential entrant is less likely to come up with an invention worth developing, so the need to do defensive research may decline.

A as long as the following sufficient condition is satisfied: $(1-p(e^{i*}(A)))>-[p'(e^{i*}(A))]^2/p''(e^{i*}(A))$.

Proof. See the appendix. ||

As long as $e^{i*}(A)$ is relatively small, the sufficient condition will be satisfied for a wide range of concave $p(e)$ functions. To take a couple of examples, the condition always holds near $e=0$ for the family of power functions $p(e)=e^{\alpha}$ with $\alpha<1$, as well as for those of the form $p(e)=e/(ae+b)$, with $a>1$. Thus as long as the incumbent's equilibrium reseach effort is not too large, one can argue that the innovation-bunching effect of Proposition 6 will dominate any potential ambiguities, so that the probability of a new product making it to market unambiguously declines with the age of the incumbent firm.

The bottom line of this subsection is therefore a simple one. As long as the incumbent does not do "too much" research in equilibrium, the previous results about waves of innovation basically carry over to the case where both incumbents and potential entrants do research. Indeed, in the limiting case where $p^{\max}$ approaches zero, the strong results of Section 3.1 apply exactly. And in the more general case where $p(e^{i*}(A))$ is significantly different from zero, but still "small" in the sense of Proposition 7, the essential flavour of the results still comes through.

4. CONCLUSIONS

The premise of this paper is a simple one: even in industries in which innovation and knowledge spillovers are critically important, firms are unlikely to compete solely on the basis of such innovation-driven variables as product quality. Rather, they will also compete in part of the basis of variables such as the strength of their respective customer bases. With customer bases there is learning-by-doing at the product level, but there are no spillovers across firms.

Introducing customer bases into a dynamic model of repeated innovation significantly alters the model's predictions. First, customer bases can in some circumstances—though this need not always be true—dramatically reduce the long-run average level of innovation. Or stated somewhat more generally, the potential for firm- and product-specific learning-by-doing can, ironically, be quite harmful to long-run growth.

Second, customer bases tend to generate endogenous waves in innovative activity of the sort described by Schumpeter (1936). The key to these waves is what I have termed a shakeup externality. When a new firm successfully enters the market, it breaks the incumbent's stranglehold on the customer base. This in turn makes it more attractive for the next generation of innovators to enter.

The model developed above is extremely stripped-down and stylized. This was done so as to illustrate the important consequences of customer bases in the simplest possible way. Unfortunately, this simplicity also makes the model less empirically realistic, and hence less appropriate for directly confronting the data. Given that some of the most interesting evidence regarding innovative industries (e.g., Gort and Klepper (1982), Klepper and Graddy (1990)) centres on the *number* of firms present at any point in time, perhaps the most glaring deficiency of the model is the very artificial industry structure in which only one firm at a time is ever active.

In principle, it should be possible to remedy this deficiency without losing either the central economic intuition of the shakeup externality or the resulting conclusion that innovations tend to come in waves. For example, the model might be extended to incorporate the idea that new entrants face increasing costs in adjusting the scale of their operations. That is, while new entrants bring with them an improved technology, they initially have limited capacity and cannot immediately start producing with this new technology on the same scale as incumbents. This would imply that incumbents are displaced only gradually by new entrants.

One appeal of extending the model in this direction is that it would link entry in any period t with an increase in the number of firms active in the market. Consequently, one should be able to generate the prediction that a major innovation that shakes up an industry would typically be followed not just by a flurry of increased innovation and entry per se, but also by a period of supra-normal growth in the number of firms in the industry. This sort of prediction could then be compared to the stylized facts on industry evolution that emerge from the work of Gort and Klepper (1982) and Klepper and Graddy (1990). For example, the shakeup externality and the accompanying waves of innovation might be helpful in explaining what Gort and Klepper (1982) call the "take-off" phase of an industry—i.e., the period during which an industry makes dramatic improvements in productivity at the same time that the number of producers is growing rapidly.[18] This would seem to be a promising avenue for future research.

APPENDIX

Proof of Lemma 1. First, I establish that the entrant immediately gains 100% of the market share in period t. To see why, suppose not, so that the incumbent has non-zero market share. For this to be an optimal strategy for the incumbent, it must make a non-zero profit in at least some period after $t-1$. This implies that the price must exceed $C_t^i(1+d\beta^{\bar{A}})$ in at least one period. But if this is the case, then by (7), the entrant can do strictly better by cutting its price in that period so that it is not an equilibrium for the incumbent to have any market share.

Second, it must be shown that the entrant charges a price of $C_t^i + D_t^i$ in period t. This follows immediately once it has been observed that the incumbent can never hope to earn any profits after t. This implies that the incumbent will never cut its price below marginal cost in period t and thus we have the standard one-shot Bertrand outcome. ||

Proof of Proposition 3. I will prove the proposition for the case where an increase in customer bases is parametrized by an increase in d. The logic for the case where it is parameterized by a decrease in β is essentially the same. A sufficient condition for an increase in d to increase the unconditional probability of innovation is that $d\lambda^*(A)/dd > 0$ for all $A \leq \bar{A}$. Differentiating (13) with respect to d, we have

$$\frac{d\lambda^*(A)}{dd} = \frac{d\lambda^*(1)}{dd}\left[\frac{(1+d\beta^A)}{(1+d\beta)}\right] + \lambda^*(1)\left[\frac{\beta^A - \beta}{(1+d\beta)^2}\right] \tag{A.1}$$

Thus a sufficient condition for $d\lambda^*(A)/dd > 0$ is:

$$\frac{d\lambda^*(1)}{dd} > \frac{[\beta - \beta^{\bar{A}}]\lambda^*(1)}{(1+d\beta)(1+d\beta^{\bar{A}})} \tag{A.2}$$

By differentiating equation (11) in the text with respect to d, we can establish that:

$$\frac{d\lambda^*(1)}{dd} = \frac{1 - \dfrac{\partial F^e}{\partial d} - \dfrac{\beta}{\lambda^*(1)}}{\dfrac{\partial F^e}{\partial \lambda^*(1)} - \dfrac{(1+d\beta)}{[\lambda^*(1)]^2}} \tag{A.3}$$

18. Other recent papers have presented different theories to rationalize the "take-off" phase of an industry's evolution. For example, in Hopenhayn (1993), growth in the number of firms is driven by exogenous growth in demand. And in Jovanovic and MacDonald (1994*b*) there is an exogenous change in the technological environment that makes it suddenly possible for firms to pursue innovative activitity.

Now consider the limiting case where $p \to 0$. In this case it can be shown that:

$$F^e \to d\left(\sum_{i=1}^{\bar{A}} \frac{(1-\beta^i)}{(1+r)^i} + \sum_{i=\bar{A}+1}^{\infty} \frac{(1-\beta^{\bar{A}})}{(1+r)^i}\right) \tag{A.4}$$

$$\frac{\partial F^e}{\partial d} \to \frac{F^e}{d} \tag{A.5}$$

$$\frac{\partial F^e}{\partial \lambda^*(1)} \to 0 \tag{A.6}$$

Using (A.3)–(A.6) and substituting in the formula (11) for $\lambda^*(1)$, the sufficient condition (A.2) can, after much simplification, be reduced to the one given in Proposition 3. ||

Proof of Proposition 6. One can write the incumbent's problem in recursive form as:

$$J(\hat{A}) = \max\left\{-e^i(\hat{A}) + [1-(1-p(e^i(\hat{A})))W(\hat{A})]\left[(d-d\beta^{\hat{A}}) + \frac{J(\widehat{A+1})}{(1+r)}\right]\right\} \tag{A.7}$$

where, for shorthand, I have defined $W(A) = p(e^{e^*}(A))G(\lambda^*(A))$. The first-order condition for this problem is:

$$p'(e^{i^*}(\hat{A})) = \frac{1}{W(\hat{A})\left[(d-d\beta^{\hat{A}}) + \frac{J(\widehat{A+1})}{(1+r)}\right]} \tag{A.8}$$

Now let us consider the specific versions of these two equations for the cases where $A=\bar{A}-1$ and $A=\bar{A}$, respectively:

$$J(\bar{A}) = \max\left\{-e^i(\bar{A}) + [1-(1-p(e^i(\bar{A})))W(\bar{A})]\left[(d-d\beta^{\bar{A}}) + \frac{J(\bar{A})}{(1+r)}\right]\right\} \tag{A.9}$$

$$J(\bar{A}-1) = \max\left\{-e^i(\bar{A}-1) + [1-(1-(e^i(\bar{A}-1)))W(\bar{A}-1)]\left[(d-d\beta^{\bar{A}-1}) + \frac{J(\bar{A})}{(1+r)}\right]\right\} \tag{A.10}$$

$$p'(e^{i^*}(\bar{A})) = \frac{1}{W(\bar{A})\left[(d-d\beta^{\bar{A}}) + \frac{J(\bar{A})}{(1+r)}\right]} \tag{A.11}$$

$$p'(e^{i^*}(\bar{A}-1)) = \frac{1}{W(\bar{A}-1)\left[(d-d\beta^{\bar{A}-1}) + \frac{J(\bar{A})}{(1+r)}\right]} \tag{A.12}$$

The first thing to prove is that $W(\bar{A}) < W(\bar{A}-1)$. To see why, suppose not so that $W(\bar{A}) \geqq W(\bar{A}-1)$. By (A.11) and (A.12), this would imply that $e^{i^*}(\bar{A}) > e^{i^*}(\bar{A}-1)$. But then from the entrant's first-order condition (18) in the text, this would imply $e^{e^*}(\bar{A}) < e^{e^*}(\bar{A}-1)$, and we would have a contradiction, along the same lines argued in the heuristic proof in the text. So $W(\bar{A}) < W(\bar{A}-1)$.

Now take this fact and apply it to (A.9) and (A.10). It follows immediately that $J(\bar{A}) > J(\bar{A}-1)$. Knowing this, we can repeat the entire argument, now comparing the case where $A=\bar{A}-1$ to that where $A=\bar{A}-2$. Using exactly the same logic, it can now be established that $W(\bar{A}-1) < W(\bar{A}-2)$, and that therefore $J(\bar{A}-1) > J(\bar{A}-2)$. This can be repeated for all values of $A < \bar{A}$, leading to the conclusion in Proposition 6, namely that $W(A)$ is a decreasing function of A.

Note also that the function $I(A)$ in the text is given by $I(\hat{A}) = (d-d\beta^{\hat{A}}) + J(\widehat{A+1})/(1+r)$. So in establishing that $J(A)$ is an increasing function of A, we have also established the same for $I(A)$. ||

Proof of Proposition 7. We want to see under what conditions the function $Z(A) \equiv (1 - p(e^{i^*}(A))W(A)$ will be decreasing with A. Differentiating, we have:

$$\frac{dZ}{dA} = (1 - p(e^{i^*}(A)))\frac{dW}{dA} - W(A)p'(e^{i^*}(A))\frac{de^{i^*}(A)}{dA} \tag{A.13}$$
$$= (1 - p(e^{i^*}(A)))\frac{dW}{dA} - \frac{1}{I(A)}\frac{de^{i^*}(A)}{dA}.$$

The first term in (A.13) is negative as desired, thanks to Proposition 6. But the second term is ambiguous in sign. The second term can be attacked by differentiating the incumbent's first-order condition (19), to obtain:

$$\frac{p''(e^{i^*}(A))}{[p'(e^{i^*}(A))]^2}\frac{de^{i^*}(A)}{dA} = -\left[I(A)\frac{dW}{dA} + W(A)\frac{dI}{dA}\right]. \tag{A.14}$$

Substituting into (A.13), we have:

$$\frac{dZ}{dA} = \left[(1 - p(e^{i^*}(A))) + \frac{[p'(e^{i^*}(A))]^2}{p''(e^{i^*}(A))}\right]\frac{dW}{dA}$$
$$+ \frac{[p'(e^{i^*}(A))]^2}{p''(e^{i^*}(A))}\left[\frac{W(A)}{I(A)}\frac{dI}{dA}\right]. \tag{A.15}$$

The last term in (A.15) will always be negative, since $dI/dA > 0$, as demonstrated in the course of proving Proposition 6. This implies that a sufficient condition for $dZ/dA < 0$ is that the term multiplying dW/dA in (A.15) be positive. This is just the sufficient condition given in Proposition 7. ||

Acknowledgements. This research is supported by MIT's International Financial Services Research Center and the National Science Foundation. Thanks to Maureen O'Donnell for help in preparing the manuscript, and to Owen Lamont for research assistance. I am also grateful for the comments and suggestions of Daron Acemoglu, Patrick Bolton, Ricardo Cabellero, Ken Froot, Rebecca Henderson, Julio Rotemberg, Alwyn Young, and the referees.

REFERENCES

AGHION, P. and HOWITT, P. (1992), "A Model of Growth through Creative Destruction", *Econometrica*, **60**, 323–351.

CABALLERO, R. J. and JAFFEE, A. B., (1993), "How High Are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth", *NBER Macroeconomics Annual 1993*, 15–74.

Continental Bank Roundtable on "Global Competition in the '90's", *Jorunal of Applied Corporate Finance*, Spring 1993, 38–59.

FAMA, E. (1985), "What's Different About Banks?", *Journal of Monetary Economics*, **15**, 29–36.

GILBERT, R. J. and NEWBERY, D. M. (1982), "Preemptive Patenting and the Persistence of Monopoly", *American Economic Review*, **72**, 514–526.

GORT, M. and KLEPPER, S. (1992), "Times Paths in the Diffusion of Product Innovations", *Economic Journal*, **92**, 630–653.

GROSSMAN, G. M. and HELPMAN, E. (1991*a*), "Quality Ladders in the Theory of Growth", *Review of Economic Studies*, 58, 43–61.

GROSSMAN, G and HELPMAN, E. (1991*b*), "Quality Ladders and Product Cycles", *Quarterly Journal of Economics*, **106**, 557–586.

HENDERSON, R. (1993), "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry", *RAND Journal of Economics*, **24**, 248–270.

HOPENHAYN, H. A. (1993), "The Shakeout" (Working Paper, Universitat Pompeu-Fabra).

JOVANOVIC, B. and MACDONALD G. M. (1994*a*), "Competitive Diffusion", *Journal of Political Economy*, **102**, 24–52.

JOVANOVIC, B. and MACDONALD, G. M. (1994*b*), "The Life-Cycle of a Competitive Industry", *Journal of Political Economy*, **102**, 322–347.

KLEPPER, S. and GRADDY, E. (1990), "The Evolution of New Industries and the Determinants of Market Structure", *RAND Journal of Economics*, **21**, 27–44.

PETERSEN, M. and RAJAN, R. (1994), "The Benefits of Lending Relationships: Evidence from Small Business Data", *The Journal of Finance*, **49**, 3–37.

PORTER, M. E. (1980) *Competitive Strategy* (New York: The Free Press).

RAJAN, R. (1992), "Insiders and Outsiders: The Choice between Informed and Arm's Length Debt", *Journal of Finance*, **47**, 1367–1400.

ROMER, P. M. (1990), "Endogenous Technological Change", *Journal of Political Economy*, **98**, S71-S102.

SCHUMPETER, J. A. (1936) *Theory of Economic Development: an Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle* (Cambridge, MA: Harvard University Press).

SCHUMPETER, J. A. (1942) *Capitalism, Socialism and Democracy* (New York, Harper and Brothers).

SEGERSTROM, P. S., ANANT, T. C. A. and DINOPOULOS, E. (1990), "A Schumpeterian Model of the Product Life Cycle", *American Economic Review*, **80**, 1077–1091.

SEGERSTROM, P. S. (1991), "Innovation, Imitation and Economic Growth", *Journal of Political Economy*, **99**, 807–827.

SHARPE, S. (1990), "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships", *Journal of Finance*, **47**, 1343–1366.

SHLEIFER, A. (1986), "Implementation Cycles", *Journal of Political Economy*, **94**, 1163–1190.

TIROLE, J. (1988) *The Theory of Industrial Organization* (Cambridge MA: MIT Press).

YOUNG, A. (1993), "Invention and Bounded Learning by Doing", *Journal of Political Economy*, **101**, 443–472.