INSTITUTIONS, INERTIA, AND CHANGING INDUSTRIAL LEADERSHIP

Paul L. Robertson
Department of Economics and Management
University College
University of New South Wales, ADFA
Campbell, ACT 2600 Australia
Internet: PLR@ccadfa.cc.adfa.oz.au

and

Richard N. Langlois
Department of Economics
The University of Connecticut
U63 Storrs, CT 06269-1063 USA
Internet: Langlois@UConnvm.UConn.edu

March 1994

Forthcoming in *Industrial and Corporate Change*

INTRODUCTION

One of the most persistent debates surrounding economic change is whether it is incremental or revolutionary in nature; whether, for instance, a period of change that lasted anywhere from seventy to one hundred years may be properly termed an Industrial *Revolution* (Landes, 1991; Cameron, 1985). Unless we abandon standard models of causality and are willing to invoke an occasional *deus ex machina* to explain change, the incrementalists are of course correct in insisting that, in the end, any event represents a change in the existing framework in the sense that it flows from (can be explained by) antecedent events and conditions. This, however, begs the question of the *rate* of change since it is still entirely conceivable that there are periods when change accelerates or slows relative to other periods. Moreover, even a moderate rate of change may be consistent with significant discontinuities in that some economic agents (individuals, firms, or other institutions such as the regulatory system) may drop out altogether and be replaced by others. In this case, seeming stability on a macro level may mask compositional effects that have a great impact on the various components of the economy.

Variations in the speed and extent of change have been noted in many fields. Gersick (1991) has found parallel theories of punctuated equilibrium in models that explain the behavior of individuals, groups, organizations, scientific fields, biological species, and the world in general.[1] When punctuated equilibrium prevails, long periods of stability and consolidation alternate with brief periods of drastic, indeed revolutionary, change in which the previous equilibrium is destroyed and a new system installed. Although Gersick (1991) does not explicitly approach product markets from this angle, her theory can also apply when, as a result of a major innovation, a particular good or service or production process is replaced by an alternative.[2]

Several features of punctuated equilibrium stand out. Firstly, it is a lengthy process. Even the revolutionary or transitionary phase, in which two or more alternatives vie for success, may be prolonged for decades, or eons in the case of speciation. Secondly, the process, like Schumpeter's "creative destruction", is one of replacement. When there is punctuated equilibrium, the extinction of a species or discrediting of a scientific theory are not enough; there must be a new species available to take over the territory or a new theory to account for the phenomena that the old theory was once thought to explain. Thirdly, each period of punctuated change requires a behavioral shift to ensure alignment between the requirements of the new order and the actions of its agents. This shift might be accomplished internally, if the old agents adapt their behavior to meet the new conditions, or externally if they are supplanted by a new group of agents. Finally, inertia plays a central role in punctuated equilibrium by ensuring that change proceeds by fits and starts rather than smoothly and evenly.

Inertia is the focus of this paper. As is explained in more detail below, inertia has two major functions in the cycle of punctuated equilibrium. Inertia results from, and in a sense embodies, the best features of the stable phase of the cycle because it is based on the learning process in which producers determine which procedures are most efficient and effective. Once people are satisfied that they know how to do things well, they have very little incentive to look for or adopt new methods. In the words of Tushman and Romanelli (1985, pp. 197, 205), "Those same social and structural factors which are associated with effective performance are also the foundations of organizational inertia. . . . success sows the seeds of extraordinary resistance to fundamental change." Inertia also provides the tension, however, that leads to the (relatively) short, sharp shock of the revolutionary period (Gould, 1983, p. 153) because the pressure required to displace a successful but inert system is considerable and takes time to accumulate. When there is little inertia, change can be assimilated in a gradual and orderly fashion, but an entrenched system may need to be vigorously displaced.

Inertia has little importance in a neoclassical world of perfect knowledge and instantaneous adjustment. Even if we relax these assumptions, the effects of inertia might be expected to be slight. In contrast to the biological world, for example, "genetic matter" in an economy is seldom truly lost: It is far easier to reinvent the wheel than to bring back dinosaurs. Furthermore, economic "species", be they technical or organizational, are routinely crossbred to create superior outcomes without much risk of sterility. Both of these factors introduce a higher degree of flexibility in markets than prevails in biology, which should mitigate the consequences of inertia.

Nevertheless, it is clear that inertia and punctuated equilibria do influence economic change in important ways. A very abbreviated list of industries in which dominant firms have been replaced by new entrants as a result of innovation includes aircraft engines, when jet propulsion took over from piston engines; transistors versus vacuum tubes; diesel-electric versus coal-fired locomotives; electronic calculators versus mechanical adding machines; and electric versus mechanical typewriters (Sahal, 1981, ch. 6; Tushman and Romanelli, 1985, pp. 200-201). More recently, innovations in process technology appear to be leading to massive displacements in industries such as automobiles.

Here we concentrate on explaining the part played by inertia in causing economic displacement. We argue that inertia is often a rational response for firms or governments even after an important innovation becomes available, and that changes in economic leadership, whether on the level of the firm or the nation, may be inevitable when there is significant innovation.

INSTITUTIONS AND INERTIA

There is a range of explanations of inertia. One set is the "real" or, in the narrow sense, "economic" explanations that look to abstract variables like demand levels, factor endowments, and relative prices to justify the failure of some organizations to change. A second reason for inertia is simple incompetence, when managers are either too stupid or too idle to adopt desirable new methods. This is a popular explanation of Britain's relative economic decline after 1870 (Aldcroft and Richardson, 1969) and is also consistent with recent comments on American businessmen attributed to Japanese leaders. Alternatively, there may be cognitive or informational problems. Managers may not have access to new knowledge or they may not recognize improvements because that do not fit their preconceptions (Gersick, 1991, p. 18).

Another set of explanations for inertia relies on cultural incompatibilities. For example, Wiener (1981) claims that the structure of British society since the end of the nineteenth century has discouraged entrepreneurship and innovation.

Here, we concentrate on the influence of institutional variables on inertia.[3] Institutions may either retard or encourage innovation. If the institutional structure is unsuited to a new technology and inert, change will be difficult to implement. When existing institutions are flexible or well-adapted to the requirements of an innovation, however, change will be accomplished relatively easily. As innovating firms may be affected by different sets of institutions, it is possible that one group may be impeded in its attempts to innovate while another group has a "head start" because it has already gained access to some of the necessary institutions for other reasons.

Both exogenous and endogenous institutions can affect the rate of innovation, as can a wide variety of institutional arrangements that are (depending on the level of analysis) either semi-endogenous or semi-exogenous. Exogenous institutions are those that are features of the economy or society at large, such as tariffs or the tax system in a given nation. Endogenous institutions comprise those that are specific to a particular firm or industry, including research and development departments, codified and uncodified corporate rules and procedures, and trade associations and lobbying groups.

In one kind of institutional explanation, the impeding or encouraging structures are only partly endogenous. That is, the institutions that retard or enhance adoption of new organizational and technological practices have been created by the self-interested behavior of individuals, but, once created, stand somehow outside or independent of that behavior. For example, the profit-maximizing behavior of a set of firms may lead to a particular kind of labor-relations policy, which in turn leads to the creation of a particular sort of labor union. Once created, this union may then affect, exogenously as it were, the subsequent behavior of the firms.

A second type of semi-endogenous institutions is those that are generated by wider societal action and may not be especially targeted at the firm or industry but which nevertheless create conditions that vitally affect the ability of the firm to innovate. Most nations, for example, have systems of schooling that turn out potential workers with an assortment of skills of various grades. Employers must then provide further training and education to meet firm-specific needs. When the system of public education is appropriate to the technological and other needs of employers, little further training is needed. If, however, there is no public education or it is directed to other ends or otherwise inadequate, a far greater burden falls on the employers. Moreover, what is highly efficient in one context may be entirely inappropriate to meet the challenges posed by significant innovations. As a result, some firms may find that a hitherto excellent system of education has become a source of inertia, while other firms, that were not well-served with educational facilities to cope with the old technology, discover that types of education that they had thought to be impractical or useless do, in fact, mesh well with the needs created by innovation.[4]

There is also a sense in which the institutions retarding change can be more straightforwardly "endogenous." The behavior patterns of the individual actors -- their routines[5] -- are themselves "institutions." Like institutions more broadly, routines (and their organizational and technological correlatives) can become obsolete. And institutional change, we argue, can often take place through the more-or-less slow dying out of obsolete institutions in a population and their replacement by better adapted institutions -- rather than by the conscious adaptation of existing institutions in the face of change.

The nature of markets, themselves, may also provide an institutional explanation for inertia because of the low level of price elasticity in many mature industries. There is little incentive for established firms to sponsor innovations if lower prices (or *de facto* lower prices as when extra longevity more than compensates for higher sticker prices) will lead to decreased total revenue as a result of inelastic demand. Innovation will be favored, however, by new entrants which stand to gain market share at the expense of existing concerns if they can exploit better product or process technologies. This helps to explain, for example, why it was Wilkinson Sword rather than Gillette that introduced stainless steel razor blades, and why American tire manufacturers were reluctant to produce the long-life steel-belted radial designs promoted by Michelin.[6]

Another set of institutional explanations that we have explored in more detail elsewhere (Robertson and Langlois, 1992) concerns firm and industry structure, in particular the extent to which a firm is vertically- or horizontally-integrated or entwined in a network relationship with suppliers, customers, or competitors. The effect of these relationships has featured importantly in a number of recent works on the position of American manufacturing industry in the world economy, with authors such as Piore and Sabel (1984), Porter (1990) and Best (1990) arguing in favor of networks or industrial districts to encourage a rapid rate of innovation while Florida and Kenney (1990) and Lazonick (1990, 1991a, 1991b) support large vertically-integrated firms.[7] In this paper, we restrict ourselves to pointing out that no single form of organization is appropriate for all, or even a majority of, cases in which innovation is desirable.

ON ROUTINES AND CAPABILITIES

Overall, then, inertia exerts two principal influences on the ability of firms to cope with innovation. Inertia is often a product of successful adaptation to earlier innovations, as a firm develops ways of operating that appear to be so well suited to its internal and external environment that it sees no reason to change. In many instances, this adaptation may prove so effective that the firm can retain a total cost advantage for a prolonged period despite using an outdated technology because it can still capitalize on its mastery of compatible support and ancillary operations, while firms adopting a new, and technically more efficient technology, are still wrestling with the expensive process of acquiring the endogenous and exogenous institutional backup necessary to gain full value from the innovation (Hannan and Freeman, 1989).

When inertia retards the learning process necessary to deal with a subsequent important innovation, however, firms that are otherwise in a position to make the eventual transition to a new technology may be so slow in coming to grips with change that dominance shifts to new entrants who are unencumbered by prior developments, learn new adaptive procedures more quickly, and are able, therefore, largely to appropriate the market by the time the established firms have learned to cope with the innovation. The obstacle in this case is may be termed "lockout", as leaders using the old technology find that they cannot successfully make the transition when there is a significant innovation (Cohen and Levinthal, 1990, p. 137).

Routines and capabilities are at the heart of both of these aspects of inertia.

From the standpoint of the internal operations of a firm, the adoption of an innovation may be conceived of as a form of diversification. The Wrigley (Scott, 1973) and Rumelt (1974) classifications of the degree of diversity or relatedness of intrafirm operations are based on the extent to which technological or marketing activities are shared across operations. The adoption of a radically new product or process technology for an existing product would lead a firm into unfamiliar, albeit still related, territory, in the same way as would diversification into a new product which shared marketing or technological bases with other products of the firm. The ability of the firm to master such a change would then depend on whether it possessed the technical and organizational flexibility to cope with an extended range of activities.

Three decades ago, in what is arguably still the richest treatment of the subject, Edith Penrose outlined a number of conditions that might induce a firm to undertake a strategy of diversification, which can be extended to cover innovation. According to Penrose (1959), firms have a tendency to acquire surplus quantities of both material and human resources. Firstly, because of indivisibilities, firms may obtain excess amounts of one or more inputs. Resources must often be purchased in bundles and, except in the relatively rare situations in which the inputs needed for the amount produced happen to coincide with the "least common multiple" of the bundles, there will be surpluses. Secondly, the efficiency of human resources tends to increase over time as personnel become more knowledgeable and more adept at dealing both with the external environment and with the administration of the firm itself.

The firm has a clear incentive to make good use of these excess resources. The way in which it employs its surpluses, however, will vary according to the firm's strengths. Not only will the varieties of resources differ among firms, but many of the most important types of resources are heterogeneous. In particular, human resources involving entrepreneurship, management, or research are not standardized, rendering each firm unique because it has kinds and combinations of resources different from those of other firms.

Penrose believes that this heterogeneity of resources will have a strong endogenous influence on the strategy adopted since each firm should attempt to make the best use of its surpluses given the qualitative nature of its own strengths. While there may be a range of uses to which a particular combination of excess resources can be put, the firm will tend to choose those that fit most closely with the types of knowledge and scope of operations that have evolved from earlier experience because these are likely to prove most profitable.

Although her terminology differs, Penrose anticipated many of the most important ideas later elaborated by other writers. In particular, Richardson introduced the useful term *capabilities* to refer to the skills, experience, and knowledge that a firm possesses. He concludes that firms "would find it expedient, for the most part, to concentrate on similar activities," that is on those that require common capabilities (1972, p. 895). Ansoff (1965) and Panzar and Willig (1981) employed a wider definition than Richardson of the attributes that firms may build on in choosing the scope of their activities. These include excess capacity in marketing, production, raw material procurement, and finance, as well as managerial or entrepreneurial knowledge, skills, and experience. For convenience, however, we will follow Teece (1980) in using "capabilities" to refer to all of these attributes.

Another aspect of capabilities that has recently received a great deal of attention is organizational culture. In practice, not all organizations may be equally able to cope with change, as existing patterns of behavior involving both executives and subordinates may be resistant to change. Organizations develop collective habits or ways of thinking that can be altered only gradually. To the extent that a given culture is either flexible or consistent with a proposed change in product or process technology, the transition to the new regime will be relatively easy. If, however, the culture is incompatible with the needs posed by the change and is inflexible, the viability of the change will be threatened (Robertson, 1990; Langlois, 1991; Camerer and Vepsalainen, 1988).

Nelson and Winter have formulated an economic analogue of capabilities, including organizational culture. "Routines," as they put it, "are the skills of an organization."[8] In the course of its development, a firm acquires a repertoire of routines that derives from its activities over the years. To the extent that these routines are efficient and difficult to come by, they are a most important asset, but they also induce inertia because they are difficult for the firm to change once in place.[9]

Teece discusses the positive aspects of such routines: That they may contribute to a capability that enables a firm to undertake new activities that are compatible with its current activities (1982). Teece neglects the negative side of Nelson and Winter's analysis, however, and fails to note that the inflexibility, or inertia, induced by routines and the capabilities that they generate can raise to prohibitive levels the cost of adopting a new technology or entering new fields. Such inertia can develop to the extent that existing rules are both hard to discard and inconsistent with types of change that might otherwise be profitable.

In adopting a product or process innovation, therefore, firms must look for a total-cost solution by weighing up possible increases in transaction costs caused by a departure from their existing capabilities and routines against savings or profitable marketing opportunities brought about by the change in technology. Moreover, the ensuing transaction costs may have two components, those that derive from disruption of existing operations and those that result from the need to learn a new set of capabilities appropriate to the new product or process technology.

Technological change, as we have shown, comes in a variety of forms that affect the likelihood of it being assimilated into existing firms. First, it is necessary to distinguish between minor and radical changes. A technological change may be characterized as "a bit-by-bit cumulative process until it is punctuated by a major advance." In general, these frequent minor changes can be assimilated in passing, and characterize the equilibrium stage. This is not true, however, of major innovations of the revolutionary stage of punctuated equilibrium, which are "advances so significant that no increase in scale, efficiency, or design can make older technologies competitive [in direct cost terms] with the new technology." (Tushman and

Anderson, 1986, p. 441.) Assuming that the adoption of a major innovation is feasible, the speed of adjustment will depend on the compatibility between the capabilities required by the old and new technological regimes. Some innovations are "competence destroying" whereas others are "competence enhancing" for particular organizations. Whereas major competence-enhancing innovations may, in time, be assimilated, the creation of entirely new organizations may be needed to deal with innovations that undermine the capabilities or competences of existing firms. Alternatively, there may be existing firms in other fields that are better able to cope with the innovation because it demands capabilities that, perhaps fortuitously, are compatible with their existing routines (Tushman and Anderson, 1986, pp. 439-65).

LEARNING AND INERTIA

Learning is the antidote to inertia because it allows organizations to switch paths by augmenting their routines and capabilities. Organizations that learn quickly, cheaply, and accurately therefore have a degree of flexibility that is denied to organizations that can only learn slowly or at great expense, or that cannot learn at all. Thus, while "[i]nertia is . . . a profoundly functional organizational characteristic in stable/predictable environments" (Tushman and Romanelli, 1985, p. 195), it is ultimately destructive when it impedes learning at times of significant change.

Stiglitz (1987) distinguishes between learning by doing and learning by learning. Under the familiar concept of learning by doing (Sahal, 1981, pp. 108-110), organizations improve their efficiency and effectiveness through experience. Stiglitz applies this notion to the learning process itself. As he explains (1987, p. 130),

> Just as experience in production increases one's productivity in producing, so experience in learning may increase one's productivity in learning. One learns to learn, at least partly in the process of learning itself. . . . By specialization in learning, one may improve one's learning skills.

But learning is not an all-embracing process. Rather, it is localized in that learning about one field of study may not yield significant increases in an organization's ability to learn about other fields (Stiglitz, 1987, pp. 126-30). There may be some spillovers that result simply from the process of learning how to question, but specific knowledge about a technology in a given industry may be of little value in dealing with a particular innovation in the same industry but with unfamiliar characteristics.[10] Recently, Stiglitz's observations on learning by learning have received some support from Kelly and Amburgey (1991, p. 606) who show that prior experience in dealing with a similar type of change increases the chances of an organization coping successfully with subsequent changes, but experience in dealing with dissimilar changes does not.

Some types of learning can also be picked up externally, by watching and benefiting from the experience of others. Learning by learning remains important, however, because some knowledge is tacit and cannot be verbalized and transferred to outsiders, while other knowledge is proprietary and not publicly-available. The ability of an organization to overcome inertia by learning is therefore limited by the timing of the learning effort and the method of learning that is chosen. Both Spence (1981) and Silverberg, Dosi, and Orsenigo (1988) have shown through simulations that lack of learning presents substantial barriers to entry. In large part, this is because organizations can readily pick up knowledge in the public domain but will be less efficient than experienced competitors if they cannot tap their own sources of tacit and proprietary knowledge. As the only way that an organization can learn the latter[11] is through learning by doing, the later the organization enters a new field or adopts an innovation, the further behind it is likely to be in efficiency. In such cases, established firms faced with mastering an innovation may be faced with barriers to entry in the same way as new entrants to the industry. They may therefore find it hard to make the transition from the old to the new technology if they delay for very long.

Cohen and Levinthal (1989, 1990) call the ability of a firm to pick up information from external sources (and thus to fend off inertia to a degree) its "absorptive capacity". They contend that an organization's absorptive capacity for external knowledge is a function of its existing knowledge. Thus organizations that already have some background in a given area may find it quicker and cheaper to acquire new related knowledge than do organizations with no prior experience in the area. Cohen and Levinthal point to basic, or generalized, R & D activities as an important way to improve a firm's chances of spreading its external nets widely in acquiring useful knowledge from its surrounding environment. But, as it is not feasible to have a basic background in all areas, the problem is still to determine which fields are likely to prove sufficiently fertile in the future to justify an investment in basic background knowledge now. Firms that do not make the correct decisions (that do not know how to learn what they specifically need to learn) may lose irrevocably. In the words of Cohen and Levinthal (1990, p. 138):

> A firm without a prior technological base in a particular field may not be able to acquire one readily if absorptive capacity is cumulative. In addition a firm may be blind to new developments in fields in which it is not investing if its updating capability is low. Accordingly . . . firms may not realize that they should be developing their absorptive capacity due to an irony associated with its valuation: the firm needs to have some absorptive capacity already to value it appropriately.

Furthermore, as Penrose (1959) noted, some organizations have better initial learning capabilities than others. Each organization is unique and its ability to acquire the knowledge necessary to adopt a significant innovation successfully differs from that of existing or potential competitors. If the innovation is competence destroying, the inertia generated by mastery of an older technology may preclude the rapid acquisition of knowledge that will permit the transition. Competence-enhancing innovations, on the other hand, can benefit either existing firms or new entrants depending on whether the competences that are strengthened are related to or distinct from those associated with the old technology.

THE POPULATION DYNAMICS OF MARKET DOMINANCE

As vital elements of internal learning are needed, first to determine which capabilities a change demands, and then to master them, inertia will be strong and the adoption of major competence- or capabilities-destroying innovations can be expected to be gradual. In the interim, the industry may be composed of two sets of firms, the representatives of the older technology who will gradually wither, and those of the new technology who, as will be shown, may or may not gain the momentum required to establish themselves permanently. The survival of the older technology rests on the mastery of appropriate capabilities by existing firms who have learned to make the most efficient use of their resources under existing conditions. If capabilities for the new technology have not yet been worked out, therefore, a prolonged period may follow in which the total cost of production of the representatives of the old technology is less than that of the newer because the transaction cost savings arising from the use of efficient routines more than offset the direct savings in production costs that can be attributed to the new technology (Hannan and Freeman, 1989, ch. 4). In fact, under certain circumstances, it may pay firms to continue to invest in a dying technology even though they would incur an accounting loss as a result (Tang, 1988).

But, within any given population of firms, the withering of the representatives of the old technology and their replacement by firms that have adopted the innovation are not symmetrical processes. This is because there may be competition between as well as within populations. It is conceivable that there may be different endowments of capabilities and other resources that make the firms in one population, for example the producers in an industry in a particular country, better able to adopt the new technology than producers elsewhere. An obvious example would be an endowment of some vital mineral that is highly localized, expensive to transport, and unnecessary under the old technological regime. If an innovation rendered this mineral necessary, not only would firms using the old technology in locations distant from the mineral deposits be at a severe cost disadvantage if they innovated, but distant new entrants would also face a severe handicap.

Equally importantly, there may be artificial differences among populations that lead to differential rates of success in adopting the new technology. This could arise, for instance, if one nation had, for independent reasons, already invested in a set a capabilities needed for the efficient use of an innovation, say those associated with technical education, but other nations had not. If this nation's firms were thus enabled to achieve rapid control over the new technology, they could potentially appropriate the innovation and gain a lasting market dominance (Abramovitz, 1986, p. 388).

A second, and more pervasive, artificial means of gaining market dominance is through the use of tariffs. Assume that, under an existing technology, production in an industry is controlled by firms in a single country who have gained an early lead. These firms have used their cost advantages, based in large part on the efficient use of learned capabilities, to blanket domestic and export markets to such an extent that there are few foreign competitors.[12] When a major new technology is developed that dramatically reduces direct costs of production, many of the existing firms are initially reluctant to adopt the innovation because their capabilities still give them an overall cost advantage. A few pioneers, however, venture into the new technology and slowly develop the capabilities necessary to use it at its most efficient level. After a period of time, the capabilities of the representatives of the old technology are no longer great enough to compensate for higher direct costs of production and these firms are obliged either to adopt the new technology belatedly (a risky and probably futile gesture[13]) or to quit the field.

In the meanwhile, a second country, which has not had a successful group of firms using the old technology, imposes a tariff (or equivalent trade barrier) to protect local firms that adopt the innovation. Because of the tariff, these firms do not have to compete against the foreign first movers whose mastery of capabilities associated with the old technology still gives them an advantage in the early stages of adoption. Furthermore, if the pioneering nation gives no tariff protection to its own producers, firms in the follower country are able to compete on equal, or nearly equal, terms with the representatives of both the old and new technologies in the pioneer's home market. If the market in the follower nation is large enough to accommodate available economies of scale, firms adopting the new technology there are able to move down their learning curves much faster than similar firms in the pioneering nation that face competition from imports as well as from the local firms that have retained the old technology. Under this scenario, it is entirely plausible that the adopters of the innovation in the follower nation are able to learn so much more quickly that they can appropriate the greater part of the market by the time the older firms in the pioneer nation finally succumb, and that the adopters of the innovation in the pioneer will have been relegated to a minor role or eliminated altogether.[14]

Nevertheless, the retention of the old technology by the pioneering firms may be rational. For example, a discounted cash flow analysis could show that the pay-off to "harvesting" the existing operation, by reducing investment and letting it run down to the point of extinction, is greater than that from shifting to a new technology because of the much higher profits to the old technology in the early years, when capabilities appropriate to the innovation are still under development. David (1991) gives a good example of harvesting in the defense of DC electrical power that Edison mounted during the Battle of the Systems. According to David, Edison was not being quixotic when he took elaborate, and sometimes bizarre, steps to contain the spread of an AC power system that even he must have known was superior in many ways. Edison needed funds for experimentation in new areas and had neither the patents nor the financial resources to enter AC transmission himself. What he really wanted to accomplish was an orderly transition that would permit him to liquidate his substantial investments in the DC technology at a good price so that he could get on with his new work. By resisting until the rotary converter was perfected, which allowed AC power to be converted to DC and thus ensured the viability of the existing DC network,[15] Edison was able to sell out at a greater profit than if he had either sold during the earlier period when the continued viability of DC was uncertain or shifted to AC as a follower.

The diagram in figure 1 illustrates graphically the possibility of harvesting and a consequent shift in leadership among firms. The solid lines are experience curves (Abell and Hammond, 1979) for two different basic technologies for producing the same good. The downward slope of the curves derives not only from learning by doing, but from such factors as economies of scale and minor competence-enhancing innovations. Indeed, the curves are drawn on a double-log scale, which emphasizes that, in the early stages, relatively small increases in cumulative output lead to relatively large decreases in production costs, but after the product matures much larger increases in cumulative output are needed to generate the same absolute decreases in production costs.

Initially, there are four firms employing the old technology: an experienced and relatively low-cost producer at A_1, an inexperienced and relatively high-cost producer at B_1, and two intermediate firms at C_1 and D_1. Firms A, C, and D are in one nation and firm B in another. Assume now that there is a major innovation that, if implemented, shifts a producer to a new and lower experience curve.[16] Both firm A and firm B have equal capabilities to adopt the innovation and would therefore be at the same point on the new experience curve (A_2, B_2). Firm B would find the innovation desirable and adopt it, but firm A would face inertia because it is already producing at far lower cost. Not only would firm A endure lower profits if it adopted the innovation, but it might even face extinction if other domestic producers using the established technology (Firms C and D) declined to innovate and instead engaged in stiff price competition. Because of its higher initial cost structure, firm B, of course, would not be able to survive during this transition phase unless it were somehow insulated from competition. This, however, is perfectly feasible if B is in a different country and can be protected by a tariff.

After a period of time (years, or perhaps even decades) in which both firms have gained experience on their respective curves, A will have moved to A_3 and B to B_3, at which points their production costs are equal. But B's costs will be falling faster. If firm A wishes to change to the new technology now, however, the best it can hope for is to enter at A_4 because, although it will have access to publicly-available knowledge concerning the innovation, A will not be able to tap the tacit and proprietary knowledge that firm B has gained. C and D will be at a similar disadvantage. Thus, the leading firms when the old technology was dominant will become the followers after the new technology takes hold because they will always have less knowledge and therefore relatively higher costs if they do not innovate initially at the same time as Firm B. Under such circumstances, if the transition period is long and the initial cost differential high, the rational course for firms A, C, and D is to harvest their investments by collecting higher profits over the period of transition even though the eventual result is to become followers or perhaps be driven from the industry when the innovative technology has become established.

The same basic mechanisms can operate either for entire national industries, as illustrated here, or for leading firms within a particular national economy that are less well-equipped than some domestic competitors to cope with a significant innovation.[17]

The evolutionary explanation presented here should be distinguished from the now-familiar notion of technological lock-in (David, 1985; Veblen, 1915, pp. 126-7). In the broadest sense, we are, of course, arguing for a kind of lock-in. Because of the continuing possibility of learning along a particular technological or organizational path, there are, in effect, transaction costs impeding movement to a new path. This is very much a matter of path-dependency. But the cause of the lock-in is not (necessarily) increasing returns arising from the presence of network externalities or from fixed costs in complementary activities. Rather, the dependence on path arises simply from the persistence of routines. In a neoclassical world of fully informed actors, one needs a specific nonconvexity to achieve lock in. In a more believable world of ignorance and bounded rationality, the following of rules -- as a necessary tool of cognition -- is enough to do the trick.

FOUR HYPOTHESES

The analysis we present suggests a class of explanations that help to predict which firms will appropriate the benefits from an innovation. Here, we illustrate four hypotheses with a variety of historical and modern examples.

Hypothesis 1. A firm that is adept at employing an existing technology will be less likely to adopt a new technology that is incompatible with its current capabilities than will a firm that is less adept at using the existing technology, even if the new technology offers the prospect of long-run increases in profits for both firms.

This is the hypothesis with which we are most concerned. It is illustrated in Britain after 1870 by electricity generation and electrical machinery and, perhaps most paradigmatically, by cotton textiles. The relatively stunted development of the electrical products industry in Britain is a reflection of an older legacy, in this case in the use of gas for lighting and of steam, pneumatic, and hydraulic tools in industry. The electrical-machinery branch of the industry was especially weak in Britain, but British firms had trouble in competing even in simple items. Imports of electric lamps in 1908 were nearly as large as total domestic production in 1907, and imports of incandescent light bulbs were half again as large as domestic production. The older telegraphic equipment branch, and in particular cable manufacturing, was somewhat stronger, but in general the British produced less sophisticated varieties of equipment for home consumption and for export to the underdeveloped areas of the globe (Byatt, 1968, 1979).

In large part, the failure of the British to attain greater success in electrical products manufacturing can be traced to the slow adoption of electricity for lighting, traction, and power purposes in the United Kingdom. Many of the reasons for this were political. By 1913, there were only a few large generating stations in both Chicago and Berlin, serving centralized power and light systems in each city. By contrast, "Greater London had sixty-five electrical utilities, seventy generating stations averaging only 5,285 kw. in capacity, forty-nine different types of supply systems, ten different frequencies, thirty-two voltage levels for transmission and twenty-four for distribution, and about seventy different methods of charging and pricing." (Hughes, 1983, p. 227.) This failure to consolidate was based on legislation that gave each municipality effective control over both public and private supply within its boundaries. Moreover, as late as 1912, the majority of electrical power used in Britain was actually generated by users rather than purchased from generating stations, further increasing the fragmentation of supply (Byatt, 1979, ch. 6; Hughes, 1983, pp. 227-38, 249-50).

In several important respects, however, the slow adoption of electrical power can be traced directly to the existing provision of other types of power in Britain which were not matched by developments in the U.S.A. or Germany. In contrast to the United States, inexpensive gas lighting was available in many British cities before electric lighting was feasible. Even in the early years of the twentieth century, electrical power was more expensive relative to gas in Britain than in the U.S.A. (Byatt, 1979, p. 24). The use of electricity for industrial purposes in Britain was retarded in two ways by existing structures. Many firms were already using steam or other forms of power, and the slow rate of expansion of the economy left less scope than in Germany or the U.S.A. for the construction of entirely new plants that could be laid out to make the best use of electrical power rather than, as in the past, built around a central steam power plant that optimized the use of shafts and belting but did not take advantage of the flexibility that electricity could provide. In addition, the structure of the economy itself reduced the spread of electrification since cotton textiles and coal mining, which were of greater importance in Britain, were less susceptible to conversion to electrical power than many fast growing industries in America and Germany (Byatt, 1979, chs. 3-5).

Nonetheless, it remains the case that the institutions retarding change in the electrical industries were in large part quasi-exogenous governmental ones. For this reason, the cotton textile industry provides a purer example in which the principal "institutions" retarding change were the endogenously developed skills and capabilities of firms and workers.

Mass and Lazonick (1990) have recently provided a thorough summary of the debate on the British cotton industry in which they have been prominent participants. Their analysis of the sources of British dominance in the industry draws in most respects on a set of facts that are not broadly in dispute, even if these authors place their own interpretation on them. The facts are these. Combining their early capabilities in the premechanical textile trades with mechanical innovativeness, the British developed a set of productive capacities in mass-produced cotton textiles ahead of any other nation. These capabilities arose in an extremely decentralized manner that partook of Marshallian external economies. Among these capabilities was a highly developed market in cotton fiber. Technologically, these capabilities came to center around the self-acting mule and the power loom, which guided incremental innovation and conditioned the skills of the labor force. Although the exact productivity figures may be in dispute, it is clear that these competences in the British economy allowed the industry to follow an experience curve like those in figure 1. And Britain was further along this curve than all others.

In the 1890s, however, a new technological paradigm emerged with the development of the automatic loom by Draper in the U.S.A. This device allowed for higher throughput and required less-skilled operatives. In order to benefit from its advantages, however, one needed yarn more resistant to breakage (for a given count) than that produced by the mule. The ring frame, which also required less labor skill and which, quite rationally, was little used in skill-rich Britain, thus became an important complement to the automatic loom, as it produced the needed stronger yarn. Coupled with new techniques for blending grades of cotton for ring spinning, the combination of the automatic loom and the ring offered a technological trajectory different from the one on which Britain had embarked. The U.S.A., Japan, and others speedily adopted these techniques behind tariff walls. In the Mass and Lazonick account,[18] Japan was particularly adept at honing the labor skills complementary to this new technology. And after World War I, that country led a pack of low-wage countries -- all of which depended on ring spinning and automatic looms -- in a successful assault on British dominance. Britain attempted belatedly to adopt the new technologies, but found itself perennially behind on the experience curve.

Hypothesis 1 is also consistent with the experiences of a number of modern firms. In addition to the razor blade and tire industries, which are discussed above, the deterioration of the position of American automobile manufacturers is also a function of inertia, in this case in the face of important changes in process technologies by Japanese firms. These changes, which to a large extent involve organizational innovations, are based on a total rethinking of the nature of mass production in the light of improvements in quality monitoring mechanisms in recent decades. By linking low costs in mass production to high quality output -- in contrast to the Taylor-Ford premise that there is a trade-off between cost and quality -- Japanese automobile manufacturers have been able simultaneously to gain a significant advantage in costs of production and a marketing advantage sustained by a reputation for high quality. Despite eventual adjustments by U.S. and European automobile producers, it seems clear that a substantial share of the market has been permanently ceded to the Japanese (Womack, Jones, and Roos, 1990; Smitka, 1991).

Hypothesis 2. If there is no major innovation in an industry, the incumbent leaders will probably retain leadership.

It is well-established that firms that gain market dominance are likely to hold onto it unless there is a major technological change (Lieberman and Montgomery, 1988). To reverse the examples cited under Hypothesis 1, the following firms all held market leadership under relatively stable conditions for several decades until challenged by major innovations: Gillette in razor blades; Goodyear, Firestone, and Uniroyal in tires; and General Motors, Ford, and Chrysler in automobiles. Similar examples of long-term dominance in mature markets apply in typewriters, electrical products, and many other industries.

In the late nineteenth and early twentieth centuries, Britain industries were also able to retain their market dominance when there were no major innovations, even when firms in related industries were being supplanted by foreign competition. British cottons maintained their dominance of international markets until well into the twentieth century, particularly in piece goods. Imports failed to penetrate the home market and, despite increased foreign competition, there was seldom any absolute decrease in exports except in the case of yarns (Tyson, 1968). In engineering the older sectors such as cotton-textile machinery also kept their international dominance. The ring frame may have been a competence-destroying innovation from the point of view of the British spinners, but it was competence-enhancing from the point of view of the machinery firms, who strengthened their position through its production (Saxonhouse and Wright, 1984). The market for British locomotives and rolling stock held up as well, especially in the Empire and South America. In boilers and prime movers, Britain exported more than the United States or Germany in 1913, and in agricultural machinery, more than Germany (Saul, 1968).

Hypothesis 3. If there is a major innovation, but it is highly compatible with existing capabilities (is competence-enhancing), the incumbents will probably retain leadership.

In their discussion of the automobile industry, Abernathy and Clark (1985) list a number of innovations that were highly significant but nevertheless competence-enhancing. Moreover, they note that, since a firm can have importance competences in several areas, an innovation that is destructive of one competence can bolster others. Thus the introduction of the inexpensive V-8 engine and steel bodies eroded some of the technical competences of major firms but allowed them to enhance their existing marketing competences.

A further example in Britain is the advent of the steel hull for oceangoing merchant vessels in the 1870s. This innovation reduced hull weights by as much as 15 per cent in comparison to iron hulls and offered improved strength and flexibility. It was, however, a change that fitted in well with the capabilities of both existing shipbuilders and steel makers in Britain. By the 1890s, virtually all major ships were built of steel and Britain had retained its leadership (Pollard and Robertson, 1979).

Hypothesis 4. If the innovation is, for all practical purposes, entirely new (that is, there is no significant existing industry), then the benefits of the innovation will be appropriated by the population of firms that already has the best access to the most important relevant capabilities. These may be either entirely new firms or incumbent firms in related fields.

EMI's experience with the CT scanner illustrates the problems a technical pioneer can encounter if it does not possess complementary capabilities (Teece, 1986). Although EMI had the technical sophistication to develop the scanner, the device required a higher standard of training, servicing, and support than hospitals had needed until then. EMI was not in a position to provide these services, but GE and Technicare, two firms that were similarly sophisticated in electronics, were also established medical suppliers. As Teece (1986, pp. 298-99) reports the final result:

> By 1978 EMI had lost market share leadership to Technicare, which was in turn quickly overtaken by GE. In October 1979, Godfrey Houndsfield of EMI shared the Nobel prize for invention of the CT scanner. Despite this honor, and the public recognition of its role in bringing this medical breakthrough to the world, the collapse of its scanner business forced EMI in the same year into the arms of a rescuer, Thorn Electrical Industries, Ltd. GE subsequently acquired what was EMI's scanner business from Thorn for what amounted to a pittance. . . Though royalties continued to flow to EMI, the company had failed to capture the lion's share of the profits generated by the innovation it had pioneered and successfully commercialized.

Mitchell (1989) has since shown that incumbent firms with related capabilities in other branches of medical diagnostic imaging are also more likely to enter new fields in imitation of pioneers than are incumbent firms without related capabilities.

Another example of Hypothesis 4 is the substitution of iron hulls and steam propulsion for wooden-hulled sailing vessels around 1850 (Pollard and Robertson, 1979). These changes thoroughly undermined the distinctive competence of North American shipbuilders, which lay in their sources of cheap timber, and instead placed a premium on access to iron plates and steam boilers, both fields in which Britain had already established the basis for leadership. Similarly, Britain developed a strong bicycle industry at the end of the nineteenth century, based in large part on the country's strong external economies in the manufacture of mechanical parts (Harrison, 1969).

In other cases, late Victorian and Edwardian Britain did not possess the right capabilities to capture the benefits of wholly new innovations. In some electrical and chemical industries, for example, countries with systems of technical education more developed than that in Britain were in a better position to appropriate the benefits of innovations that relied on the use of technically-trained labor. As in the case of other types of education, it was difficult to align the costs and returns from technical training and, as the government and potential students were less willing to pay in Britain than in some other countries, technical education there was stunted (Robertson, 1981). Firms in Britain, which could rely on their initial capabilities arising from a stock of skilled labor, were confirmed in their reliance on manual skills and became less able as technologies evolved to make the transition to more highly technical routines. This, in turn, determined not only the work practices of firms in older industries but helped to steer British entrepreneurs away from industries where technical knowledge was vital because they knew that suitable workers would be expensive to hire. In Germany, the U.S.A., and Switzerland, a readily expandable pool of technically-educated workers was already available when the electrical and chemical industries began to take hold. As these foreign firms had many of the necessary capabilities for innovation on hand before the innovations were actually adopted, they were able to learn other routines much more quickly and to seize market share before British firms could become meaningful competitors.

CONCLUSIONS

Institutional factors, especially those embodied in capabilities and routines, can both improve the ability of a firm to exploit an existing technology and make it more difficult to innovate by generating an inertia that is hard to overcome. As a result, periods of technological change are often relatively short and dynamic in comparison to lengthy periods of consolidation in which firms gain full mastery over innovations. Not all organizations are equally well equipped to adapt to change, however, and firms that are adept at using an existing technology may have fewer of the capabilities required to cope with innovation than a new entrant or a firm that was less successful under the old regime. When this is true, a change in industrial leadership is probable, with the hitherto dominant firms becoming either followers or leaving the industry altogether because they are no longer competitive.

A number of policy implications flow from this analysis:

1. It may be highly rational for a firm to cling to an old technology, even if it has a limited lifespan. The return to harvesting may be greater than that to innovating for a firm that has strong routines and capabilities relevant to the old technology but has fewer capabilities suited to the innovation than other firms.

2. Governments should be wary, however, of propping up firms that do not have the necessary capabilities to cope with change. Too little attention is sometimes given to the "destruction" aspect of creative destruction. It can be very expensive to help firms to cling indefinitely to outdated technologies or to pay them to acquire capabilities that other firms have gained more cheaply. To the extent that some firms are encouraged to persist with obsolete technologies longer than they otherwise would, the adoption of an innovation by other firms with better capabilities may be retarded, causing a long-term cost to society.

3. If it is nevertheless felt, either by private firms or governments, that they need to obtain a foothold in the innovative technology despite a high degree of rational inertia, it is best to begin adjusting as soon as possible. Otherwise, competitors may have acquired so much experience

with the innovation that late adopters will be hard pressed to catch up in the acquisition of tacit and proprietary knowledge.

4. One way of handling innovation when a firm has good reasons to remain inert is through "tapered adoption". In this way, through pilot projects it will be possible to acquire knowledge and avoid falling too far behind despite the probable losses that the innovation will bring in the earlier stages. The experiments can then be gradually expanded to replace the old technology as the cost advantage shifts towards the innovation. If an industry is believed to be of great strategic or economic importance, governments may wish to encourage firms to embrace tapered adoption.

5. The analysis also offers evidence that infant industry arguments can make sense if a nation whose firms have been followers under an old technology believes that there are sufficient capabilities available to support an innovation. In such cases, there could be a substantial long-run payoff to providing tariff protection for domestic innovators so that they can develop capabilities while inertia encourages overseas competitors to continue to use the old technology.

NOTES

[1] Among the authors that Gersick cites for these various fields are (1) *Individuals* - Levinson (1978, 1986); (2) *Groups* - Gersick (1988, 1989); (3) *Organizations* - Tushman and Romanelli (1985); (4) *Scientific Fields* - Kuhn (1970): (5) *Biological Species* - Eldridge and Gould (1972); Gould (1983, chs. 17 and 18; 1989); Wake, Roth, and Wake (1983); and (6) *Grand Theory* - Prigonine and Stengers (1984); Haken (1981).

[2] Tushman and Romanelli (1985, pp. 197-201) do consider the effects of innovation on product classes, but primarily from the perspective of organizational behavior rather than economics.

[3] Although general cultural variables such as "national character" are also institutions in the sense meant here, we do not consider them because of the difficulties in establishing criteria and finding useful data.

[4] What we term semi-endogenous institutions are "endogenous" in much the same way that the evolution of government institutions is endogenous to the market in the Theory of Public Choice (Olson, 1982).

[5] We develop this notion of routines presently. But on the affinity between social institutions (such as norms) and organizational routines, see the essays in Langlois (1986).

[6] Although Michelin was an established tire manufacturer in Europe, it had an insignificant share of the U.S. market and, like a new entrant to the field in general, had little to lose if the innovation were rejected in America but a great deal to gain if it were successful.

[7] For a historical perspective that favors vertical integration, see also Chandler (1977 and 1990).

[8] Nelson and Winter (1982, p. 124). Note that *routines* refer to what an organization actually does, while *capabilities* also include what it may do if its resources are reallocated. Thus a firm's routines are a subset of its capabilities that influence but do not fully determine what the firm is competent to achieve.

[9] Abramovitz makes a similar point at the level of the national economy (1986, pp. 402-5).

[10] From our standpoint, Stiglitz seems to reverse the time dimension involved in change. He writes, for example, that "the basic concept of `weaving' is involved in virtually all textile production, but much of the technical knowledge associated with modern automated factory production is inapplicable to hand-loom weaving." (1987, p. 127). It seems far more relevant, however, that the knowledge of the hand-loom weavers was not readily transferable to later technologies.

[11] This is particularly true of tacit knowledge since proprietary knowledge may be obtained through industrial espionage or other surreptitious methods.

[12] Although somewhat exaggerated, this is a fair description of Britain's position following the Napoleonic Wars (Landes, 1969, ch. 3).

[13] Cooper and Schendel (1976), Foster (1986), and Tushman and Anderson (1986) all indicate that the ability of existing firms to adjust to radical product or process innovation is highly limited.

[14] Tariffs are only one means that a late mover can use to assist its firms in achieving control over a major innovation. Other tools include regulations that make it difficult for foreign firms to become established in the follower nation, as is alleged to happen in Japan, the use of "safety" regulations to discriminate against imports, and campaigns to encourage local customers, especially governments, to buy locally-produced goods.

[15] DC power was still supplied in sections of major cities at least as late as the 1960s.

[16] In fact, an alternative definition of a major innovation is one that requires a shift to a different experience curve since minor innovations are among the factors that contribute to a movement down a particular curve.

[17] In the initial stages, before the necessary capabilities to deal with the innovation have been developed and total costs are higher for innovators than for users of the older technology, preferential government policies would probably not be available for innovators in the same economy as the current leaders under the earlier technological regime. Other devices to "protect" the innovators would be possible, however, such as the cross-subsidization of innovating divisions by other parts of diversified firms.

[18] For alternative views, see Sandberg (1974) and, in particular, Saxonhouse and Wright (1984, 1987).