THE INDUSTRY–TECHNOLOGY LIFE CYCLE: AN INTEGRATING META-MODEL?

Robert U. Ayres
International Institute for Applied Systems Analysis
Laxenburg, Austria

RR-87-3
March 1987

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

International Standard Book Number 3-7045-0081-x

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Cover design by Martin Schobel

Printed by Novographic, Vienna, Austria

Summary

This report attempts to answer a rather deep question: To what extent can "pure" economics explain economic growth and technological change? By "pure" economics, it is meant the relationships governing the behavior of abstract entities, producing abstract products or services for sale in an idealized competitive market. Pure economics, in the above sense, admits R&D and innovations of an unspecified kind; it also admits improvements (unspecified) and learning curves or experience curves.

The report concludes, however, that the dynamic behavior of the product "life cycle" in specific cases, and the observed clustering of innovations in particular fields at particular times, with periods of rapid progress followed by slowdowns, can only by explained by also taking into account the preexisting state of technology, and the laws of nature. It is argued that technological progress is marked not only by processes of relatively predictable incremental improvement (e.g., "learning curves"), but also by a series of discrete "barriers" and "breakthroughs". These are not random events, although their exact timing is undoubtedly very difficult to predict.

The technological life cycle can be defined as the period from a major breakthrough opening up a new territory for exploitation to the next major barrier. It is characterized, in part, by a high initial marginal productivity of R&D, and a more or less continuous decline thereafter, as the territory is gradually exhausted. This model is qualitatively consistent with the well-known "S-shaped curve" phenomenon, describing measures of technological performance over time, as well as a number of other observed phenomena.

Foreword

Two strands of argument are interwoven in this report. The first strand is that technological innovation is a major driver of economic growth. The second strand is that technological innovation is also a consequence of economic activity. In the jargon of the profession, technological change is essentially *endogenous*, not exogenous (as has often been assumed for the sake of convenience). In short, technological innovation and economic growth are related like the chicken and the egg: to give either priority over the other is futile.

Of course, economists have known this for some time. But more is needed. It is not quite enough to postulate microeconomic mechanisms to explain why entrepreneurs invest in R&D. We also need to understand better the characteristic technology life cycle and its relation with the better-known product or industry life cycle. These are major themes of the TES program at IIASA. Professor Ayres addresses the issue squarely in this report and suggests some directions for future research.

THOMAS H. LEE
Program Leader,
Technology–Economy–Society
International Institute for Applied Systems Analysis

Preface

This paper was begun in late 1984, when the author was in the Department of Engineering and Public Policy at Carnegie–Mellon University in Pittsburgh. It was completed in the winter of 1986, at IIASA, under the auspices of the Technology–Economy–Society Program. Since the subject is central to the TES program, it is being issued as an IIASA Research Report.

The author wishes to acknowledge constructive comments on various drafts from Wesley Cohen, Edwin Mansfield, Richard Nelson, Gerhard Rosegger, and Nathan Rosenberg. Thanks are particularly due to Harvey Brooks for his very detailed and helpful critique – almost a minipaper in itself, for which I am most grateful. Any oversimplification, neglect of nuance, or outright error which remains is entirely my own responsibility.

ROBERT U. AYRES
Deputy Program Leader,
Technology–Economy–Society
Project Leader,
Flexible Manufacturing
International Institute for Applied Systems Analysis

Contents

THE INDUSTRY–TECHNOLOGY LIFE CYCLE: AN INTEGRATING META-MODEL?

Robert U. Ayres

Background

The life cycle is a cluster of related phenomena that the economic theory of technological change should explain better. Used in economics, "life cycle" is, of course, a metaphor of succession taken from biology. The stages include conception, birth, infancy, childhood, adolescence, maturity, senescence, and death [1]. Recognizably similar stages can easily be identified in the evolution of other entities, such as a technology [2], a product [3], a firm, an industry, or perhaps even a nation. In fact, without stretching the metaphor unduly, a series of correspondences can be identified between the stages of the life cycles for technologies, products, and firms or industries. A stylized description of the life cycle follows. It is summarized in *Tables 1-3*.

The life cycle of an industry begins with a major product innovation or technological breakthrough. Schumpeter saw the innovation as a major entrepreneurial act, preceded by invention and driven by the lure of supernormal monopoly profits. Schumpeter (1912) originally treated the creation of scientific knowledge and invention as exogenous to the economic system [4]. In his later work, Schumpeter (1943) modified this view and allowed for the creation of technology (R&D) as a deliberate activity, especially in large firms. For a recent discussion of the two Schumpeterian models, see Freeman (1982).

The innovative product may or may not be immediately successful in the marketplace (*infancy*); but if it is successful, it quickly spawns both improvements and imitators. The second stage of the cycle (*childhood*), also described implicitly by Schumpeter, can be characterized as "imitative innovation". At this time there is typically an intense competition among entrepreneurs for market niches, based primarily on design and cost-effectiveness improvements, over the initial entrant. Inventors and innovators try to protect their technologies through patents and secrecy, but diffusion of knowledge occurs inevitably. Only the largest and most dominant firms can expect to capture more than a small proportion of the total benefits of an invention. In this connection, there is a distinct difference between product and process invention. The latter is easier to protect, especially after firms have grown large. However, further innovation during this period of rapid flux is sometimes motivated by a perceived need to invent around a set of dominant patents [5]. Frequently, the competition

Table 1. A summary of the life cycle: technology.

Life cycle stage	*Diversity versus standard*	*Scale and process technology*	*Labor require-ments*	*Logistics, transport, inventory, handling, technology*
INFANCY	Unique	Custom; *ad hoc* multi-purpose machines	High labor intensity, high-skilll workers needed	Little concern w/ inventory, low relative cost of trans-port, manual handling
CHILDHOOD	Diversity of types & imitators	Small batch job shop, manual operation	Expansion of labor force, but minimal "deskilling"	Inventory costs increase sharply, transport cost increase, manual handling
ADOLESCENCE	Increasing standardiza-tion, fewer models; faster diffusion	Medium to large batch; special machines & fixtures	Embodiment of labor skills in machines	High inventory, high transport costs, semi-auto handling
MATURITY	High degree of standardi-zation, approaching saturation	"Mass"; dedicated mechani-zation, transfer lines etc.	Low labor intensity, low skill for direct mfg. jobs; High skill for indirect & managerial jobs	Reduced inventory, high transport, mechanized handling
SENESCENCE	Commodity-like			

begins even earlier at the R&D stage, as a number of would-be innovators simultaneously seek to develop a new product for a widely recognized potential market. Entry to the new industry is still easy, in this stage, though many would-be entrants fail.

The transition from the second stage (*childhood*) to the third stage (*adolescence*) can probably best be characterized as the beginning of consolidation, i.e., the period when the number of entrants to the industry per year is first exceeded by the number of departures. At this period, expansion is most rapid.

Eventually, scale economies and the accumulation of "experience" [6] or "learning by doing" achieved by the most successful survivors begins to raise barriers to new entrants. This characterizes the third stage of the cycle, *adolescence*. The minimum investment required grows larger and the requisite "know-how" resides in a smaller and smaller number of existing organizations and their key employees. As time passes, quantum (nonincremental) product improvements seem to become harder to achieve. There are declining returns for R&D, and the basis for competition among the survivors shifts toward price-cutting and marketing. Eventually, a "shakeout" is likely to occur [7]. The need to

Table 2. A summary of the life cycle: strategic management.

Life cycle Stage	*Investment strategy*	*Capital intensity*	*Location*	*Organization*
INFANCY	Invest in the product	Low capital intensity maximum risk premium	Near technical talent pool or source of financial backing	Flexible, nonhierarchic; technical skills on top
CHILDHOOD	Emphasize product R&D	Growing investment in mfg. facilities; declining risk	Near primary market(s); many warehouses and variants	Still flexible; mfg. skills on top; marketing skills increasingly important
ADOLESCENCE	Expand markets, cut costs, exploit economies of scale; shift to process R&D	Increasing investment in mfg. facilities; declining risk	Transition to larger, centralized facilities; long distribution channels	Bureaucratizing; organizational and marketing skills on top
MATURITY	Emphasize process R&D; diversity	High capital intensity, minimum risk premium	Based on lowest-cost labor (or capital); moves "offshore"	Bureaucratic, financial, and legal skills on top
SENESCENCE	Disinvest: sell assets including technology to low-cost competitors	Maximum capital intensity	Loses national identity	Bureaucratic, financial and legal skills on top

Table 3. A summary of the life cycle: market and industry structure.

Life stage	*Price elasticity*	*Entry/exit*	*Competition*
INFANCY	Not applicable	Easy entry	Natural monopoly of innovator
CHILDHOOD	High elasticity	Many new entrants	Rapid imitation; increasing competition
ADOLESCENCE	Declining elasticity	"Shakeout" mergers, entry is hard, more exits	Transition and consolidation of survivors
MATURITY	Low elasticity	No new entry	Stable oligopoly
SENESCENCE	Negligible elasticity	More mergers and failures	Unstable; original leaders begin to drop out

exploit economies of scale in manufacturing and marketing encourages consolidation of niches and product standardization. This permits larger-scale production and further increased productivity (Verdoorn's law), which in turn permits lower prices. Given nonzero price elasticity of demand, this in turn translates into larger markets, a mechanism emphasized by Salter (1960). During this stage, the industry may enjoy its maximum rate of expansion.

From the technological point of view, the dominant process during this stage is diffusion [8], although the connotation of technological or entrepreneurial passivity on the part of the producer is misleading. The role of demand pull is clearly an important element of the diffusion process, as are the invention and innovation processes. However, the producer's technological response to market signals and his investment in marketing can, in many cases, strongly influence the diffusion process.

The fourth major stage (*maturity*) is reached when markets are becoming saturated and the price elasticity of demand falls toward unity, or even below. During this stage, the product becomes increasingly standardized. In late maturity (or senescence), it tends to become a commodity. Product innovation tends to be slow and planned, or nonexistent, for a standardized product. Thus, supernormal profits are possible at this stage only to the extent that markets are incontestable, i.e., barriers to market entry permit monopoly pricing (Baumol, 1982). In fact, by commonplace observation, many (but not all) technologically mature industries tend to be dominated by a few large firms. To the extent that highly capital-intensive production methods are adopted, both product and process innovation are less attractive than earlier. [The problem arises from the well-known inflexibility of traditional mass production technology, which makes both product and process innovation forbiddingly expensive (Abernathy, 1978).] This constitutes an effective entry barrier to small or new firms.

A further implication of the life cycle, noted by Vernon (1966) and others, is that production capacity for mature products tends to move from high-wage to low-wage countries, as shown in *Figure 1*. This is the most straightforward explanation of the so-called Leontief paradox in trade theory [9]. The migration of manufacturing industry from the USA to third world countries such as Korea, Malaysia, the Philippines, Brazil, and Mexico is clearly evident in recent years. The "maturity hypothesis", as applied to the economy as a whole, has been advanced particularly by Kindleberger (1953, 1962, and 1979). A major contribution in the international context is Olson's institutional sclerosis theory. Olson (1982) emphasizes the role of extended periods of political stability in permitting the accumulation of powerful and protected special interest lobbies and legislation that interfere with market mechanisms for resource allocation and thus reduce economic efficiency.

While many aspects of the life cycle have been elucidated, some gaps remain. In particular, the "aging" process and the apparent slowdown in innovation in mature industries, as contrasted with a higher rate of innovation in "adolescent" industries, are not yet explained adequately by either neoclassical equilibrium approaches or by dynamic disequilibrium models in the Schumpeterian tradition.

At this point, the so-called "evolutionary" model (or family of models) elaborated by Richard Nelson and Sidney Winter (1982) can be regarded as a promising alternative to neoclassical theories of economic growth. However, the issue particularly addressed by Nelson and Winter up to now has been whether their evolutionary nonequilibrium, nonoptimizing model can account for more or less continuous economic growth at the macro level as well as (or better than) traditional neoclassical equilibrium models. Their critique of orthodox theory is primarily directed at the interface between pure (neoclassical) economics and human and organizational behavior.

The critique offered in this paper is primarily directed, by contrast, at the other end of the spectrum: the interface of orthodox economics with the laws of nature (i.e., with the subject matter of engineering and science). It is in no way inconsistent with the Nelson–Winter point of view, except in minor respects, but may be regarded as complementary to it.

In summary, existing economic theories do not appear to offer fully satisfactory microeconomic explanations of several key phenomena:

(1) The occurrence of major breakthroughs [10] from time to time.
(2) The cluster of related innovations and improvements that tends to follow a major breakthrough.
(3) The subsequent maturation and aging process, with its characteristic slowdown in the rate of technological progress, product standardization, and so on.

(4) The possible relationship of the above to "long waves" or Kondratieff cycles. This phenomenon has been the subject of recurring debates among economists since 1913. There has been a recent revival of interest, occasioned by the present worldwide recession. Various theories have been presented to explain long waves, but the subject is highly controversial, to say the least. I will return to this topic later.

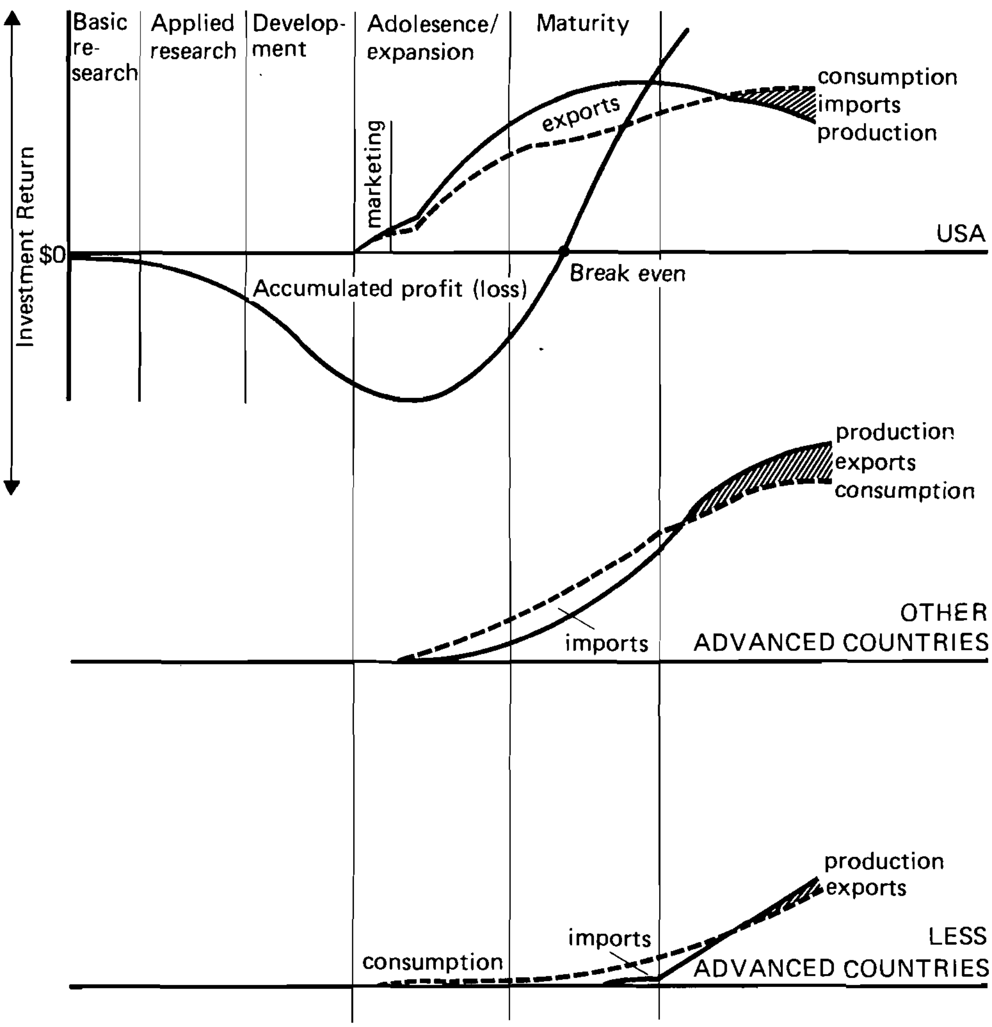

Figure 1. The product life cycle (Ayres, 1984, adapted from Vernon, 1966).

Technological Breakthroughs

As noted above, there is a widespread view among mainstream economists that the specific features of technological change are essentially unpredictable, except in the statistical sense that investment in R&D can be expected to generate useful new ideas. The contemporary orthodox view is reasonably well summarized by Heertje: "Technical knowledge, being the product of a production process in which scarce resources are allocated, can be produced. *We do not know exactly what will be produced, but we are certain that we will know more after an unknown period*" (Heertje, 1983: p. 48; emphasis added).

Scientists and engineers tend to be less pessimistic about the possibility of forecasting what will be produced by R&D. In fact, most R&D is explicitly directed toward specific ends. There are many documented instances of simultaneous invention [see, e.g., Ogburn and Thomas (1922)], which strongly suggests parallel independent searches, motivated by a widely shared perception of technological opportunity. The most famous such coincidence was that of Alexander Graham Bell and Elisha Gray, who filed documents with the U.S. Patent Office on the same day (February 14, 1876) claiming invention of the telephone. Bell worked in Boston; Gray, in Chicago. The point is that the invention of the telephone (among others) was certainly anticipated by many individuals who were familiar with telegraph technology before it actually occurred [11]. The same is true of the majority of other important inventions.

The Nelson–Winter model of technological progress is consistent with Heertje's view quoted above, viz, it assumes (for convenience) that the probability of a successful innovation is a function of R&D investment and is more or less independent of past history or other factors. Were it really so, technological progress would be much smoother than it actually is. In the real world, it is clear that opportunities within any field vary enormously from one time to another. Similarly, opportunities vary greatly from one field to another at any given time. Venture capital seeks identifiable opportunities, which can be defined as areas where a relatively small R&D effort may have a large economic payoff. It will be argued hereafter that such technological opportunities exist, from time to time, because of the changing state of technologies with respect to each other or vis-à-vis discontinuities in the laws of nature and the properties of matter. It follows, then, that many major opportunities are foreseeable, in principle, by technologically sophisticated persons. Indeed, the great majority have been foreseen in fact.

To explain this "discontinuity" hypothesis, it is useful to review, briefly, the technological history of the past two centuries. The major loci of inventive activity in the late eighteenth and early nineteenth century were iron-making, steam engines, railroads, textile manufacturing, and metal-working machine tools. In the second half of the nineteenth century steel, electrification, internal combustion engines, and chemicals took center stage. Automobiles and aircraft followed as active areas of invention, followed by electronics (radio and TV) and polymer chemistry. After World War II came spinoffs such as nuclear technology, antibiotics, jet engines and rockets, the transistor, integrated circuits, and the digital computer. Today composite materials, artificial intelligence, and biotechnology are the hot areas.

Apart from the civilian applications of military technology, the pattern that characterized development in each of these major technologies (and numerous lesser ones) is roughly the same. Initially, a few visionary scientists and/or inventors worked for many years – even decades – to overcome a long-recognized barrier before the first successful prototype of the invention could be demonstrated. In most cases, there were a number of false starts. The history of flight is typical. Even after the first successful prototype built by the Wright brothers, many more years elapsed before a product was finally ready for the "market" [12]. The initial barrier to progress may have been lack of knowledge, as in the early development of metallurgy, chemistry, and applications of electricity and radio. Around the turn of the nineteenth century, barriers tended to be the dearth of sufficiently strong or hard materials, inability to achieve desired levels of energy-density or power-to-weight, or the lack of sufficiently precise forming tools or means of measurement. In recent decades, the barrier has frequently been the lack of sufficiently fast, accurate, or reliable methods of manufacturing and/or information processing. Once the operative constraint is overcome, progress can be rapid – until the next technological barrier is reached.

One of the famous historical examples of such a constraint was the problem of accurate navigation. Scientists in the seventeenth century, including Newton, recognized that this required accurate means of timekeeping that would – unlike available weight or pendulum-driven grandfather clocks – be unaffected by a ship's irregular motion. The ultimate technical solution to the problem was the spring-wound chronometer (or watch). Unfortunately, the requisite material (good-quality spring steel) could not easily be manufactured until the advent of Huntsman's crucible steel process in the 1740s. Huntsman's process opened up a variety of collateral economic possibilities in other areas as well.

Newcomen's first steam engines overcame a constraint on the depth of coal mines. Watt's improved steam engine overcame a constraint on the availability of factory sites with access to water power. His engines, in turn, could not have been commercially produced without prior improvements in iron-making and iron-casting technology on the one hand, and Wilkinson's cylinder-boring machine (1774) on the other. Stevenson's first steam railroad locomotive (1818) could not have succeeded in practice until compact high-pressure engines became feasible, thanks to Trevithick, Woolf, and others around 1802–1804. Trevithick tried, but failed, to make a viable locomotive for mine haulage. His engine was too heavy, either for wooden rails or for the brittle cast iron rails available at the time. Practical railroads depended on heavy-duty rolled iron rails (developed by Birkinshaw in 1821), as well as steel springs for the suspension, to facilitate adhesion of the driving wheels to the track.

Engineering developments in the nineteenth century were continuously limited by the slow rate of progress in ferrous metallurgy. Textile machinery, clocks, guns, sewing machines, and agricultural machinery could not be produced in quantity without accurate steam- or water-powered machine tools for reliable cutting and forming of brass or iron parts. Such tools, themselves, depended on the quality of metal and the prior availability of adequate machine tools and measuring devices. (Thus, the early evolution of the machine tool industry was, to a degree, technologically self-limiting.) The later success of the automobile could not have occurred without the prior existence of a sophisticated machine tool industry as well as a sophisticated iron and steel industry. Above all, the automobile (and the aircraft) depended on the prior development of a light-weight reliable propulsion system, the internal combustion engine (ICE) [13].

The growth of the steel industry after 1860 also perfectly exemplifies the overcoming of a major barrier originating in the laws of nature. The smelting of iron ore to pig iron [14] requires only maximum furnace temperatures of the order of 1100°C. Such temperatures have been achievable with the help of power-driven bellows since the late Middle Ages. The pig iron is brittle and not useful as an engineering material without further refining to eliminate impurities and excess carbon. However, to convert tonnage quantities of pig iron to pure malleable (bar) iron, or to low-carbon steel, temperatures in excess of 1540°C are needed. This was impossible to achieve until 1740, and then only in very small quantities (by Huntsman's crucible method). Henry Bessemer first solved the temperature problem in the 1850s by blowing air rapidly through molten pig iron in a "converter". Kelly, in the USA, found the same solution to the problem at about the same time. It happens that oxygen in the air combines preferentially with the carbon and other impurities in the iron, resulting in a spectacular burst of fireworks, but leaving pure molten iron. The heat of rapid combustion of the carbon also raises the temperature of the iron to the needed level of 1540°C or more. (Soon afterward, Siemens and Martin introduced a slower but more controllable method of achieving the temperatures needed to refine steel, the so-called "open hearth" process.)

The marginal benefits of increasing furnace temperatures by 50°C from, say, 1300°C to 1350°C were comparatively slight, because no major new industrial capability was created thereby. Thus, when the "state of the art" was 1300°C, moderate improvements were not worth much. On the other hand, the marginal value of increasing furnace temperatures from 1500°C to 1550°C was enormous, because the latter temperature was the key to large-scale steel-making. It was the perception of this opportunity that inspired Bessemer, Kelly, Siemens, Martin, and others to undertake the necessary R&D.

It is interesting to note that further incremental improvements, say from 1550°C to 1600°C, would have been worth much less. But, once the great Bessemer steel breakthrough had been made, a host of collateral inventions became possible. To begin with, these included variants and improvements on the steel production process itself. The most important of these was the Thomas–Gilchrist process for making steel from pig iron with a high phosphorus content. There followed many applications of the "new" engineering material that had formerly been scarce and expensive. Later developments attributable to the availability of steel included large steel structures – such as suspension bridges and skyscrapers – steel wheels for railroad cars, armor plate, barbed wire, "tin" cans, galvanized roofing sheets, and mass-produced automobiles, to name but a few.

If Bessemer steel was "almost the greatest invention" [15], the first practical steam-electric generating plant and distribution network built by Edison in 1882 must have been the greatest breakthrough of all. The availability of electricity in large quantities, beginning in the 1890s, made possible the widespread use of safe and convenient electric light [16], plus electric household appliances such as water heaters, washing machines, and irons. It also permitted electric drive for trams and street railways and for industrial machinery (replacing the nineteenth-century system whereby machines received power via belts from a single shaft driven by a steam engine). The rapidly spreading availability of electricity was also a major spur to the development of telecommunication and electronics technology. Each of the above applications of electric power led to the creation of major industrial branches and hundreds of thousands – ultimately millions – of jobs.

Two of the less obvious but equally significant results of the electric power breakthrough are worthy of mention, namely, electric arc furnaces and electrolytic cells. Heroult's electric arc furnace (c. 1900) was a vital prerequisite to further progress in metallurgy, since many important steel and other alloying elements melt at temperatures higher than iron. Electric furnaces were capable of reaching temperatures above 2000°C for the first time in human history. The list of products that became possible as a direct result of this achievement includes calcium carbide (acetylene), carborundum, silicon carbide and other refractory ceramics, so-called ferroalloys, tungsten carbide, and most of the superalloys and exotic metals now used for gas turbine engines, rocket motors, and the like.

Electrolytic cells, in turn, were needed for commercial production of metallic sodium, chlorine, and aluminum, as well as for the refining of pure copper and the electroplating industry. Without aluminum, in turn, airplanes would never have been practical, and the aerospace industry could not exist today. In other words, a viable aircraft industry became economically feasible only after aluminum made from electricity became commercially available in large amounts. Similarly, the large-scale availability of acetylene (from calcium carbide made in electric furnaces) and chlorine made by electrolysis led to a plethora of important developments in chemical technology. To cite only one example, polyvinyl chloride (PVC) – still one of the most important plastics – is a polymer of vinyl chloride, which was originally derived from acetylene and chlorine. In fact, much of modern industrial chemistry can be traced back directly or indirectly to the commercial introduction of acetylene and chlorine in the 1890s.

Breakthroughs presuppose barriers. Whenever it can be said that A is(was) a prerequisite for B, it is also not unlikely that lack of A is(was) a barrier to the achievement of B. The existence of such barriers is inherent in the laws of nature and the properties of matter. A comprehensive historical survey of past technological barriers and breakthroughs would be a major undertaking and out of place here. The point that needs to be made is that, contrary to the assumption in most economic treatments of the subject, ***technological progress is not a simple or linear function of R&D invested.*** In short, the probability of a *major* innovation depends very critically on the state of the supporting technology at the time [17]. While a scientific or technical breakthrough is likely to be anticipated by many experts in the field, it is also likely to be unexpected by those outside the field in question. Yet the implications may cross many boundaries. Hence, economic opportunities are greatest, and the cost of innovation is lowest, just after a scientific or technological breakthrough, especially one that has not been widely anticipated. Conversely, opportunities decline and costs rise as a new barrier is closely approached.

The Nature of Barriers

A brief discussion of the nature of technological barriers is in order at this point. Generally speaking, they can be characterized as (i) peculiarities of the landscape, as when a hidden discontinuity (or chasm) only reveals itself as one approaches across a seemingly flat plain; or (ii) as distortions of perspective, as when a distant mountain appears much closer and smaller than it really is. The history of nuclear fusion research illustrates the former problem perfectly. The fusion-based H-bomb followed so closely after the fission-based U235 or P239 bombs that most scientists at first assumed that *controlled* fusion reactors would not be much more difficult to achieve than controlled fission reactors. It was not so, however. Further research soon revealed the chasm: a class of previously unknown magneto-hydrodynamic instabilities that would plague any known magnetic containment system for a high-temperature plasma. Progress in fusion research continues, gradually and slowly, but the earliest date of a practical fusion power plant retreats year by year. It is now clear that this date must be several decades into the twenty-first century.

Supersonic flight offers another example. It was once assumed that aircraft flight speeds would increase more or less smoothly as engine power was increased. Not so. Again, a discontinuity was found. Near the speed of sound ("Mach 1") turbulence increases sharply, and the power required to exceed sonic speed rises in a sharply nonlinear fashion. Thus, the supersonic Concorde uses several times as much fuel per passenger-km as its subsonic rivals, and the deluxe Concorde service operates at a loss, even with premium prices and massive government subsidies for the production of the planes.

Most barriers, however, are not due to special unrecognized problems – like the hidden chasm – but to simple inability to judge technological "distance" accurately. For instance, Charles Babbage sharply underestimated the cost and time required to build his famous mechanical "difference engine" from metal parts, simply because he was a mathematician with no practical engineering knowledge. Often, as in Babbage's case, it is an ancillary technology that is inadequate [18]. The importance of steel springs to navigation has been mentioned. For steel-making it was the ancillary technology of insulation and heat retention – thermal engineering – that was inadequate. The problem of manufacturing "interchangeable parts" – a recognized military goal as early as 1717 that was not fully achieved until the 1880s – was that, until then, the composition of iron and steel alloys could not be controlled precisely enough to make high-production machine tools accurate enough [19].

The self-powered road vehicle (automobile) and its cousin, the self-powered flying machine, were also anticipated long ago. Like the calculating engine, they turned out to be much more difficult to achieve than the early visionaries realized. The primary problem in both cases was lack of a motive power source that was sufficiently light and compact. A secondary (but not trivial) problem was to build a frame or body that was light enough yet strong enough to withstand the considerable stresses involved in use. Unfortunately, until the end of the nineteenth century (and even later), nobody knew how to calculate either the dynamic forces or the response of various body or frame members. Progress was therefore severely inhibited by lack of scientific knowledge. Another key mechanical component – the ball bearing – also first became available in the 1880s, and was quickly applied to the bicycle. In fact, bicycle technology was of critical importance, and it was no accident that the Wright brothers were bicycle builders. In any case, the engineering difficulties were far greater than could be imagined by even the most sophisticated physicist circa 1850.

Underestimation of difficulties has characterized computer applications from the start. One of the first research projects undertaken by IBM computer scientists in the mid-1950s was a program to translate English to Russian and vice versa [20]. Needless to say, although research continues, no satisfactory translation program has yet been unveiled to the public.

Another classic example of underestimation appears to be Herbert Simon's celebrated prediction in the 1960s that computers with "artificial intelligence" (AI) would be capable of defeating the best human chess players in 10 years. Simon was perhaps assuming a higher level of research intensity to develop AI than was forthcoming initially, but investment in the field has been growing rapidly for years and is now at a significant level by any standard. A large number of computers in various institutions have been programmed to play chess, and the former Soviet world champion Mikhail Botvinnik now spends much of his time working with one such group. Nevertheless, the best chess-playing computers are still significantly inferior to the best human players, and the progress of computers in this field has slowed. Nobody is predicting any more how long it will take for AI to overcome the gap.

A more mundane example is also illuminating. A barrier to truly efficient airline operation was the unavailability, until the 1970s, of an adequate computerized passenger reservation system. Such systems had been under development since the late 1950s, but the inherent difficulty of the task was grossly underestimated, and the first several versions were so plagued by breakdowns and errors that they were almost more trouble than benefit to the airlines. On the other hand, when such systems finally did become operational in the 1970s, they gave enormous competitive advantages to the airlines that successfully developed and owned them (American Airlines and United Airlines).

The use of computers to control manufacturing is another task whose difficulty was massively underestimated by early enthusiasts such as Norbert Wiener (1948) and John Diebold (1952). In the introduction to his famous book *Cybernetics*, Wiener speculated:

> The automatic factory, the assembly line without human agents, are only so far ahead of us as is limited by our willingness to put such a degree of effort as was set, for example, in the development of radar in the second world war.

In fact, for controlling some continuous processes, such as hot rolling or petroleum refining, Wiener's optimistic assessment was not grossly in error. However, nearly four decades later, the computerized control of discrete parts manufacturing and assembly operations has been partially realized in only a few showpiece factories, mainly in Japan. Every such factory is still unique.

Other examples of underestimation of difficulties (or overestimation of opportunities) have been discussed in detail by Freeman (1982).

Just occasionally, however, a technological problem turns out to be easier to solve than the "experts" think. This may happen because the problem had not been clearly articulated, except perhaps by a few people. Once in a while, the solution also turns out to be unexpectedly easy, or it may exist already in another field. This probably explains the few but interesting cases of "a solution seeking a problem".

The transistor and the integrated circuit perhaps best exemplify this situation. The problem Bell Telephone Laboratories set out to solve, around 1940, was that demand for telephone switching equipment (electromechanical relays) was growing so fast that it would predictably consume huge amounts of electric power by the 1960s. This was a problem, however, visible mainly to a few senior executives in the Bell system. Could a simple low-power switching device be found?

It turned out that the phenomenon of impurity semi-conduction, known in certain transitional metals, offered an effective approach. By modern standards, the search was relatively short and cheap; a three-man team at Bell Laboratories did most of the work. By 1948, transistors had been successfully fabricated and tested in circuits, and means of manufacturing had been developed. Instead of exploiting the technology in house, however, the technology was made available to licensees [21]. The monolithic integrated circuit, announced in 1958–1959, was also developed rather quickly and inexpensively at Texas Instruments and Fairchild Semiconductor in response to a need to simplify the assembly of complex circuits from large numbers of discrete components. What few people understood at first was that these solid-state devices would also soon solve another problem that was only beginning to be recognized by 1950: the inherent unreliability of electronic devices utilizing vacuum tubes. Integrated circuits quickly revolutionized all of electronics, most of all the electronic computer [22].

An Expanding Frontiers Model of the Life Cycle

In view of the points made about breakthroughs and barriers in the last two sections, a reinterpretation of the product life cycle seems called for. The sort of major innovation that initiates such a cycle often occurs because of a technological breakthrough, often in another field. Such a break-through typically makes available a new engineering material, a new sensing or control capability, greater power, speed, temperature or pressure, a new type of energy-conversion device, or a new manufacturing process. One is almost irresistibly drawn to a familiar metaphor: the westward expansion of the USA during the nineteenth century. Each time a political or topographical barrier was overcome (the Appalachians, the French-Indian Alliance, the Mississippi, the Rockies, the warlike Sioux and Blackfeet tribes, the Mexicans), colonists poured into and occupied the newly opened territories. In so doing, they created the jumping off conditions for the next phase of expansion.

I have argued above that most major clusters of innovations follow the conquest of a significant technological barrier. The territory beyond such a barrier is little known ("terra incognita"), at first, because either the means or the motives for exploring it are lacking. But, once the barrier is surmounted, all is changed: a new territory suddenly opens for exploration and dominion. The rush to claim ownership over virgin technological territory is closely analogous to the land rushes (or gold rushes) of the nineteenth century [23]. Land was claimed by homesteaders or miners under rules set by the federal government. Technological territory is normally claimed by patents, also under rules set by government.

The detailed similarities – and differences – between a gold rush or a major oil strike and a "technology rush" need not concern us unduly at this stage. The important feature of both processes is that the economic payoff per unit of exploratory (R&D) effort is likely to increase, at first, as the number of colonizers increase, because they can create and share a common knowledge base or "infrastructure". In the case of a new agricultural or mineral territory, the infrastructure means access to transportation, processing, and customers. In the case of a new technology, it means the accumulation of common basic knowledge of measuring or fabricating techniques, for instance.

Later, however, the economic payoff per unit of R&D effort is likely to decline sharply as the best available locations (or concepts) are preempted. Thus, the probability of improving significantly on the performance-level state of the art tends to decline as the state of the art itself advances. In economic language, the marginal product of R&D in any established field tends to decrease. In mundane terms, the cost of an increment of progress in a technology increases with the level of the technology.

The increasing cost of achieving decreasing degrees of improvements in product or process technology is a recognized fact of life in the R&D laboratory. Anecdotal evidence abounds, and few R&D managers would dispute the point. Unfortunately, quantitative data are surprisingly scarce. This is partly because expenditures on industrial R&D are seldom published and partly because it is very difficult to find examples in which expenditures can be allocated to improvements along a well-defined axis or measure of performance. Perhaps the best available data are from a classic study of petroleum refining (Enos, 1962).

Prior to 1913, the only fuel available for use by automobiles and trucks was natural gasoline, or naphtha, obtained by a distillation process developed in the 1870s. Naphtha is defined in terms of its "boiling" range: 75–350°F. The naphtha fraction varied with the source of crude oil, but was typically around 25%. Higher-range boiling fractions – illuminating oils (350–600°F), and lubricating oils, waxes, and residual oils (600–1000°F) – constituted around 66% of the total mass and had a very low market value compared to motor gasoline.

To increase the yield of motor gasoline, it was necessary to "crack" the heavier fractions. The first cracking process, developed by Burton in 1913, used heat and pressure as the cracking agents and doubled the gasoline yield per barrel of crude oil (Enos, 1962: p. 23). It also raised the average research octane number (RON) from about 50 for natural gasoline to 55. This advantage was not recognized, at first, but later became important when it was realized that higher octanes would permit higher engine compression ratios and, consequently, greater fuel economy.

Processes introduced since Burton's original thermal cracking development have increased the maximum gasoline yield per barrel of oxide to around 75%: the U.S. average is around 55%. But research octane number (RON) has increased dramatically, from 55 to 95. As shown in *Table 4*, the improved processes introduced in 1922, 1936, and 1942 successively resulted in performance improvements over the previous process of 21.7%, 35.3%, and 25.3%, respectively. There was no further performance improvement after 1942.

Cracking process innovations also resulted in reduced raw material, labor, and process energy costs, which were initially translated into profits for producers (return on investment) and later into savings for gasoline consumers. It is interesting to note that the new processes each resulted in significant improvements in the product (gasoline) but usually at some initial penalty in labor, capital or energy productivity. These penalties were invariably eliminated quickly by subsequent process improvements.

The major point illustrated by *Table 4* is, of course, that the cost of each incremental improvement, whether in product or in process, rose dramatically. The R&D cost of the original Burton cracking process was only $92,000. Successive new processes cost $7 million, $11 million and, $15 million, respectively, to reach the introduction stage. Dramatic improvements in factor productivity were later made in each process, but the R&D costs were also much higher for the later processes. Since 1955, refinery processes have been improved, but only to a minor extent, and further major process innovations seem quite unlikely.

Table 4. Marginal returns to R&D on cracked gasoline. Source: data from Enos (1962).

Comparisons	*R&D costs \$(000)* (a)	*Percentage improvement in productivity of:* *Gasoline* (b)	*Labor* (c)	*Capital* (d)	*Process Energy* (e)
Improved Burton process (1922; scale factor x 3) v. "original" Burton process (1913)	144	N.A.	+623%	+200%	+87%
New continuous thermal processes (1922; scale 570 bbl/day) v. improved Burton processes (1922)	7,000 (a)	+21.7%	-28%	+50%	+25%
Improved continuous process (1938; scale factor x 12) v. original continuous process (1922)	>2,000	0	+277%	240%	+38%
New Houdry catalytic process (1938; scale 6750 bbl/day) v. improved continuous process (1938)	11,000	+35.3%	-32%	-55%	+216%
Improved Houdry (1940; scale factor x 2) v. original Houdry process (1938)	1,150	0	+380%	+57%	-37%
New fluid process (1942; scale 2750 bbl/day) v. improved Houdry process (1940)	11,000	+25.3%	-45%	-12%	-42%
Improved fluid process (1955; scale factor x 3) v. original fluid process (1942)	>15,000	0	+225%	+60%	+333%

(a) Including R&D costs for both Dubbs process and "Tube & Tank" process prior to 1922 (Enos, 1962: Table 6, p. 238).
(b) Measured in terms of ton-miles at 40 mph, to take into account higher engine compression ratios permitted by higher octaves (Enos, 1962: Table A, p. 271).
(c) Productivity measured as gallons of cracked gasoline per manhour process labor.
(d) Productivity measured as gallons of cracked gasoline per dollar of capital (1939\$).
(e) Productivity measured as gallons of cracked gasoline per million BTU.

A recently published book by management consultant Richard N. Foster (1986) documents two other cases of declining productivity of R&D in quantitative terms. *Figure 2* shows a measure of two chemical processes as a function of cumulative R&D man-hours. As it happens, the two processes were direct competitors, insofar as naphthalene and orthoxylene (O-xylene) are alternative feedstocks for an important industrial chemical, phthalic anhydride. Evidently, the naphthalene-based technology was initially more efficient. But at the same time the newer orthoxylene-based technology was improving more rapidly, with a greater fractional improvement per unit of R&D investment, mainly because its theoretical maximum yield (of phthalic anhydride) was higher. Foster argues that the "defenders" in this case (Allied and Monsanto) should have deemphasized naphthalene at an earlier stage and switched their R&D to orthoxylene or simply licensed the new technology (as Monsanto eventually did).

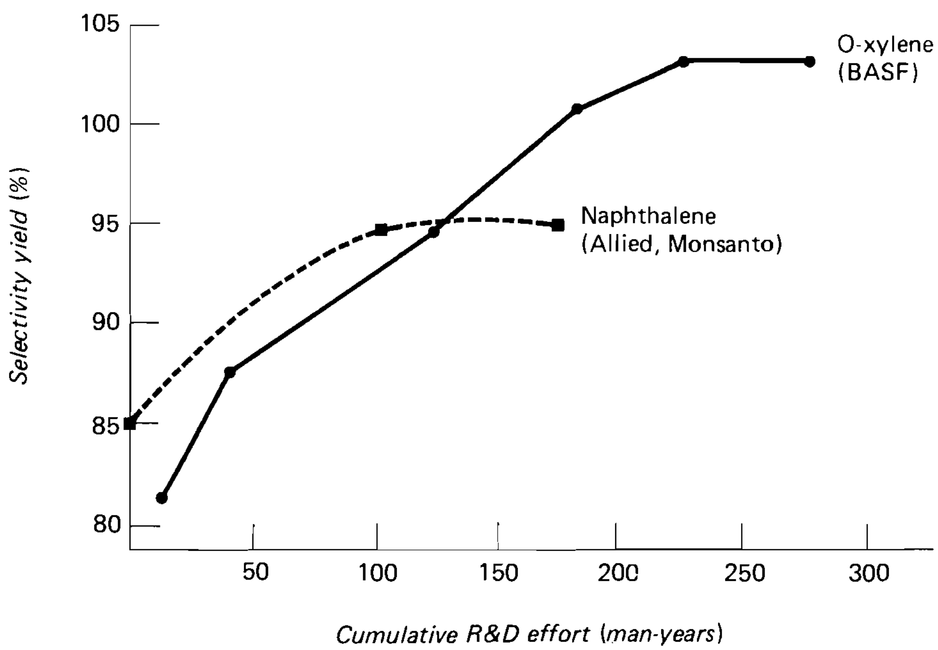

Figure 2. R&D productivity for two industrial chemicals: O-xylene and naphthalene (Foster, 1986).

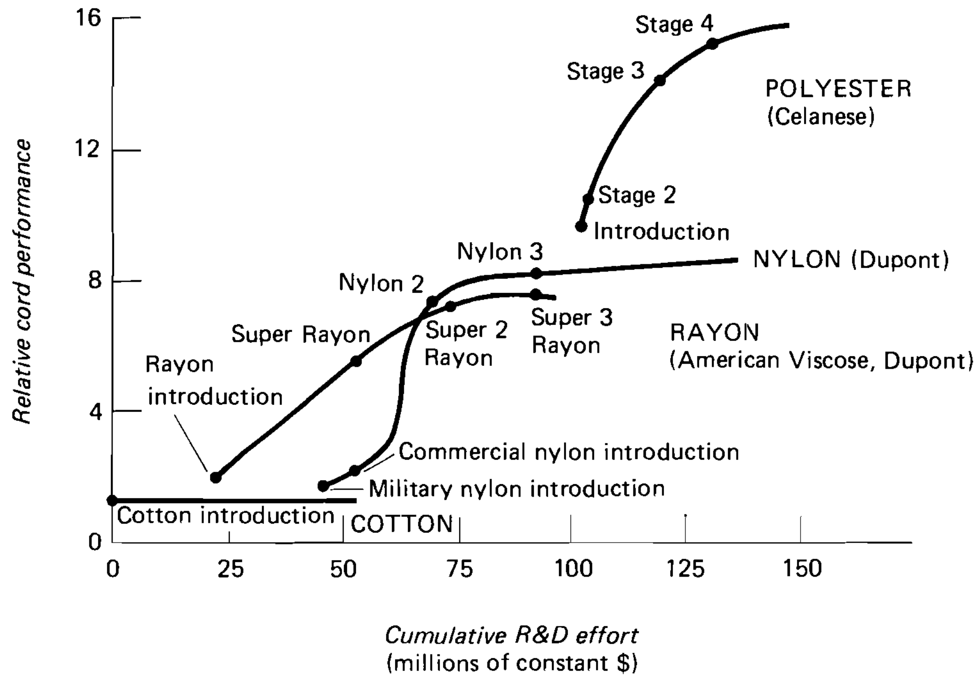

Figure 3. R&D productivity for tire cords: cotton, rayon, nylon, and polyester (Foster, 1986).

An even more interesting case of multiple substitution is shown in *Figure 3*, which shows the returns to R&D for rayon, nylon, and polyester tire cords measured against a common index of performance. Again, the decline in R&D yield for each technology is clearly evident. In fact, the first $60 million invested in rayon tire cord technology before 1962 resulted in an 800% increase in performance. The next $15 million resulted in a much more modest (but still significant) 25% improvement, while the last $25 million gained a mere 5% in performance. By that time nylon was already superior, but also rapidly approaching its limits (apparently unknown to Dupont), whereas the "adolescent" polyester technology was rapidly improving. In fact, Foster (1986) cites many cases where a younger, more dynamic technology has an R&D yield 5 to 20 times greater than an established competitor approaching its inherent limits. For this reason, picking the right technology to develop (in terms of its life cycle) is of critical importance to competitive success.

Mechanisms

Processes that accelerate at the outset and later decelerate, can be explained in terms of positive and negative feedback processes. During the acceleration stage, the feedback is positive. Later, the feedback becomes negative and the process decelerates. Examples of such processes in nature include autocatalytic chemical processes and biological growth; see, e.g., Lotka (1956: Chapter III) and von Bertalanffy (1957). Derek de Solla Price (1963) was one of the first to recognize the closeness of the analogy between such natural processes and scientific progress. Similar ideas were developed by Lenz (1962) and by Hartman and Isenson (cited in Jantsch, 1967: Chapter 11.3). The latter two authors each derived detailed models for forecasting the rate of change of the "state of knowledge" as a function of time, in terms of such variables as the number of investigators in a field, the "productivity factor" for a scientist, and a "reaction cross-section" to reflect the probability that an encounter between two scientists working in the field will generate new knowledge.

Both the Isenson and Hartman models predict S-shaped curves for the growth of knowledge in a newly opened territory. Simple approximations in both cases lead to the well-known "logistic" equation [24] for x:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = kx(y - x)$$

where x is any measure of knowledge (on technological performance), k is a growth constant and y is an upper limit, presumably based on interest physical factors. This equation is integrable and its solution is well known, viz,

$$x(t) = \{1 + \exp[-k(t - t_0)]\}^{-1}$$

The logistic function above is a special case, in that it is perfectly symmetrical around the time t_0, which is also the point of inflection of the S-curve. Nevertheless, the simple equation is widely used for explaining the growth and substitution of new technologies for old ones [25] as shown, for instance, by Mansfield (1961), Fisher and Pry (1971), Blackman (1972 and 1974), and Marchetti (1977).

Unfortunately, the diffusion of innovations is often not a symmetrical process in time: sometimes the process of diffusion slows down; in other cases, it speeds up (Gold *et al.*, 1970; Mahajan and Wind, 1985). A number of alternative innovation-diffusion models have been introduced since the late 1960s by, e.g., Floyd (1968), Bass (1969), Sharif and Kabir (1976), and Easingwood *et al.* (1983); but none is perfectly general. Also, the more general models, e.g., by Sharif and Kabir and by Easingwood *et al.*, are not integrable, except numerically (Mahajan and Peterson, 1985). This is a severe limitation for the practitioner, since non-integrable equations are very difficult to fit. The difficulty has been partially relieved, recently, by Skiadas (1985), who has introduced two integrable models that are capable of reflecting almost any S-type curve. Recent work by Peschel and Mende (1986) suggests that a general class of "hyperlogistic" equations may be appropriate. Except for the Bass model, however, none of the generalized innovation-diffusion models can claim to be derived from underlying microeconomic or behavioral principles. They are essentially phenomenological in nature.

However, from the larger perspective adopted here, the correct specification of an innovation-diffusion model is much less important than the fact that the S-shaped curve, in one form or another, is a fairly good description of the phenomena of innovation and diffusion as they actually occur. The most universal characteristic of the phenomenon is growth followed by maturity: growth is an accelerating rate of change in the technological measure due to positive feedback. Maturity is a stage defined by a decelerating rate of change, owing to negative feedback as the technology measure approaches a plateau. This occurs because the "new territory" is fully occupied and its potentialities are exhausted.

In short, the dynamic "frontiers" model clearly suggests that the phenomena of maturity and senescence associated with the industry life cycle can best be explained in terms of declining technological opportunity. Putting it another way, the model suggests that the major reason for slow observed technical change nowadays, in such established technologies as steel-making, glass-making, steam engines, internal combustion engines, power transmissions, hydraulic turbines, machine tools (except for controls) and so forth, is that the last major breakthroughs in those technologies occurred many decades ago. In each case, *a major barrier now stands in the way of further progress.* In most cases, it is probably due to practical limits in the strength of engineering materials (especially at elevated temperatures). Significant progress in materials science could thus set off a chain reaction of collateral advances in other technologies. In some other cases, the effective barrier may be sheer complexity and what has been called "the tyranny of numbers" [26].

Much more could be said about the reasons for a slowdown in innovation toward the end of a life cycle. One early economic discussion of the topic is by Brozen (1951). One reason has been discussed in the specific context of the auto industry by Abernathy (1978), and in a more general context by the present author (Ayres, 1984); it can be summarized in Abernathy's phrase: the "productivity dilemma". In brief, as an industry matures it begins to compete more and more in terms of price. Standardization of products and domination of the industry by a few large producers permit the use of very specialized, dedicated capital equipment. This, in turn, implies that any change in the product necessitates a major capital writeoff, which is treated as a loss by present-day accountants. The bigger the writeoff, therefore, the less the motivation for making changes.

It is important to emphasize that the "frontiers" model is not necessarily incompatible with conventional economic theory, although it involves exogenous elements. What is suggested, here, is a modification of the standard theory insofar as it treats R&D as a search for *unspecified* product or process improvements. The proposed model would also take into account the existing state of process and product technology and of the market. Market pull and technological opportunity are surely complementary. Both must exist, but they need not be equally important. Often the pull is strong, but the opportunities are minimal; the result is cosmetic change, or none at all. Once in a while, however, the opportunity is clear, but the pull is unfocused and hard to recognize. This is the tougher challenge, by far, both for entrepreneurs and theorists.

Technological Opportunity

A natural definition of technological opportunity emerges from the previous discussion. In brief, *a period of great opportunity exists when and where a small incremental improvement in some technology would suffice to surmount a major long-standing barrier.* The opportunities do not necessarily exist in the same field as the barrier. If they did, opportunities would be much easier to recognize. The essential point is that breaking through a barrier almost always creates new *collateral* opportunities. The most critical question for an entrepreneur is: how can potential opportunities be recognized in advance (i.e., ahead of the competition)?

The assessment of technological opportunity, in practice, involves three observable factors: (1) the performance improvement factor vis-à-vis a known barrier, (2) the rate of progress that can be expected in the near term, and (3) the identification of collateral benefits. The first two observable factors can be assessed only by technical experts, while the third lies in a kind of intermediate domain. It may be postulated that entrepreneurs can, and do, estimate their potential gains by using these three kinds of information in conjunction with economic knowledge about the market (i.e., the demand curve) and of their own costs (i.e., the supply curve). But, in practice, this knowledge tends to be quite imprecise and harder to quantify than most planners admit [27].

To summarize: it is suggested here that important innovations occur in clusters, after a breakthrough that opens up a new, unexplored territory. It is hard to doubt that innovations are made by entrepreneurs seeking supernormal profits from technology-based monopoly. However, it is not necessary to suppose that any sort of formal utility (or profit) maximization is involved. In fact, the decision process governing entrepreneurs' specific R&D investments remains largely unexplained. How does an entrepreneur decide between long-term risky investments in the hope of a major breakthrough versus short-term, less risky investments in improvements to existing products or processes? What is the optimum level of investment? What is the optimum pace for expanding production of a new product? What is the best price policy for a new product?

The most productive working hypothesis may be that technological entrepreneurs in a new field tend to seek opportunities where a perceived demand is matched to a comparative advantage on the supply side. In other words, the technological entrepreneur is an opportunist, hoping to cash in on some preexisting knowledge (i.e., comparative advantage) that is not yet widely diffused.

In the evolution of a field of technology, the most important discoveries and the most valuable inventions tend to occur early, while by far the greatest number of individual advances occur much later, after the field is established. Indeed, it is almost always possible for the entrepreneur to know *ex ante* whether a successful R&D outcome would be a minor or major improvement or a breakthrough. One question yet to be answered is what information the entrepreneur uses to arrive at this judgment.

In cases such as biotechnology today, where a successful R&D outcome will predictably result in a large (i.e., order of magnitude) improvement in either the performance of some biological product or its cost, in terms of service yield per dollar, the entrepreneur will be correspondingly more strongly motivated to succeed. *Ceteris paribus*, he will invest more in the search, to cover as many possibilities as he can, as fast as he can. Of course, any such opportunity is likely to be known to many others, so in principle the entrepreneur will face greater competition – possibly from larger and better-financed rivals. This certainly increases the risk and reduces the odds of any clear-cut victory in the race. The essential point is that entrepreneurs generally do have a reasonably good idea of the value of the potential opportunity, *ex ante*.

It will be argued elsewhere that firms can have different effective time preferences (discount rates) in different circumstances on the basis of external threats and/or opportunities [28]. It would seem equally plausible that aggregate technological investment behavior should vary over time for similar reasons. At times of relatively low external threat (i.e., steadily expanding markets for existing products; no excess supply), the effective internal discount rate is positive, and there is no reason to take big risks on radical new technologies. They are likely to remain on the shelf in such periods. However, at times of depression, market saturation, and/or rapid expansion of capacity by low-cost suppliers, the effective internal discount rate may well be small enough (or negative) to justify risky investments with long-term payoffs.

Economic Implications of the Life Cycle Model

I have reviewed some of the economic and technological evidence supporting a dynamic expanding frontiers model of the technology life cycle. It is now appropriate to reverse the argument and ask: supposing the proposed dynamic model to be "valid" (whatever that may mean), what would it imply in terms of the major debates in the literature? Three topics are of particular interest:

(1) The classic technology-push versus market-pull argument implicitly rests on the existence of a static equilibrium. In the life cycle of a technology, the balance between push and pull changes over time. In the very early period, technology push may be quite important. In some extreme cases, "Say's law" is applicable: supply creates its own demand. This was almost certainly true of X-rays, penicillin, nylon, DDT, lasers, and genetic engineering, to name six examples. None were expected or explicitly sought in advance. They arose out of fundamental research programs yet found practical applications almost immediately.

Later in the life cycle, pull takes over. Its function is to induce a collection of competing entrepreneurs to find an optimum balance between product performance and price for the customer vis-à-vis profitability for the producer. The relative importance of pull over time may be measured roughly in terms of price elasticity of aggregate demand. The more mature an industry, the lower the price elasticity, and the smaller the potential for further market expansion. However, the cross-elasticity of demand for any one firm's product increases, which is to say that the product becomes more "commodity-like".

In summary, the importance of push is likely to be highest at the very beginning of the life cycle. As the initial innovator-monopolist is challenged by many imitators, however, pull forces become dominant. Later still, in the mature phase, the effect of pull also declines.

(2) The classic argument as to whether technology is an exogenous or an endogenous factor in economic growth is also predicated on a static equilibrium picture. Even Schmookler (1966), who perhaps gathered the most evidence in support of the endogenous view, acknowledged the existence of an irreducible exogenous element. (Six examples of unexpected and somewhat surprising innovations were listed above.) On the other hand, most discoveries and inventions are not really surprising, except to observers lacking scientific and technological knowledge.

There is nothing at all surprising about the existence of an exogenous component, once it is acknowledged that the rate and direction of technological progress depends, in part, on the state of science and technology itself. Since scientific knowledge is (by definition) never complete, there is always some chance of a surprise coming out of the laboratory. By the same token, knowledge of human reactions – hence of marketplace responses – is also incomplete. Thus, surprises can also occur on the demand side [29].

In any case, I have argued in this paper that the conventional demand-side interpretation of R&D investment behavior must be complemented by a supply-side analysis of *technological opportunity*, which is an explicit function of the current state, and rate of change, of science and technology *per se*.

(3) One of the most interesting of the current debates among economists concerns the relationship of technological change and the so-called Kondratieff long cycle or long wave, first noted in terms of wholesale commodity prices. The phenomenon was originally discovered by van Gelderen (1913) and later analyzed by a number of other, mainly Dutch economists. However, the classic studies were those of the Soviet economist N.D. Kondratieff (1926 and 1928). In his analysis of business cycles, Schumpeter (1939) tried to explain the long cycle in terms of "heroic" technological innovations – notably, steam power (1818–1842), electrification (1882–1930), and automobiles (1898–1940). Schumpeter's theory was immediately challenged by Kuznets (1940), who asked two cogent questions:

(a) Is there any evidence of Kondratieff long waves in important indicators of general economic activity?
(b) Is there any evidence of a bunching of Schumpeter's heroic innovations (and, if yes, what is the theoretical explanation)?

Kuznets' answer to these questions was "no", and he remained a skeptic almost forty years later (Kuznets, 1978). Nevertheless, the "long wave" has been reconsidered in recent years by a number of authors. For a useful review of writings by Rostow, Forrester, Mandel, and van Duijn, *inter alia*, see Kleinknecht (1986: Chapter 1).

In particular, Schumpeter's notion that major technological innovations may drive the long cycle has been taken up and carried further by Mensch (1975). The centerpiece of Mensch's "metamorphosis theory" is changing investment behavior depending on market conditions. Mensch argues that during periods of prosperity, when markets are rapidly growing, capital can be reinvested with a high return and little risk in straightforward capacity expansion. On the other hand, when existing markets are saturated, the most profitable opportunities for capital are offered by investment in new technologies, which have meanwhile been accumulating on the shelf. Mensch believes that major depressions have occurred because the marketplace is too slow to withdraw capital from mature or post-mature sectors (such as steel) and shift it to faster growing sectors.

It is evident that the dynamic life cycle model of technological change, presented in sketch form in this paper, is consistent with Mensch's ideas about investment behavior. It could help explain the long cycle if the long cycle is "real". While major peaks and valleys in economic activity on a roughly 50–60-year time scale have been observed, their statistical significance is still unclear. However, given the occurrence of a major economic depression for any reason, Mensch's theory implies that this should be a peak in the rate of major innovation. Conversely, during a period of general prosperity, the theory implies that the rate of major innovation should be low. In this connection, see Ayres and Mori (1987).

It is worth noting that this expectation seems to conflict with Schmookler's (1966) empirical evidence on the correlation of patent activity and economic activity. However, as pointed out earlier, the conflict is not necessarily irreconcilable, because Schmookler's data *ipso facto* related primarily to the behavior of already established industries and did not attempt to measure the clustering of major innovations in relation to general economic conditions. On the clustering question, some supporting evidence was presented by Mensch, although it has been criticized sharply by Freeman *et al.* (1982), Mansfield (1983), and others. Meanwhile, additional data has been gathered by Freeman *et al.*, Kleinknecht (1986), and others. The jury is still out, although it seems increasingly clear that some degree of clustering has occurred. What is not yet clear is whether the observed clustering phenomenon is accidental or whether the "causes" are essentially economic or essentially technological or political/military. Moreover, it is unclear whether the apparent clustering of major innovations is causally linked to the Kondratieff cycle (Rosenberg and Frischtak, 1984). On all these matters, more research is needed.

Notes

[1] The idea of an "aging" process goes back, in part at least, to the German economist Wolff, whose ideas were cited by, among others, Kuznets (1930). "Wolff's law" asserted that the cost of incremental improvement increases as a technology approaches its long-run performance level. A number of other economists of the 1930s explored the process of industrial succession and displacement, especially in the context of business cycles, including Burns, Hansen, Hoffman, and Schumpeter (see Schumpeter, 1939). A general "law of industrial growth" was proposed (Alderfer and Michl, 1942). In summary, it stated that industries mature when technological progress slows down, resulting in slower cost reductions and market expansion.

[2] The first explicit reference to a technology life cycle I am aware of appears in an unpublished report by Ralph Lenz (1962).

[3] The notion of a product life cycle seems to have originated with Nelson (1962); also see Levitt (1965). It was elaborated in a classic paper on international trade and investment implications by Vernon (1966).

[4] The exogenous versus endogenous debate is one of the oldest in economics. A major early contribution was made by Hicks (1932), who attributed "induced" technological changes to factor price differentials. Thus, he explained a bias toward labor-saving innovation in the USA, as contrasted to capital- or resource-saving innovation in Europe, in terms of differences in wage rates versus capital costs. There has been a great deal of research on price-induced innovation in recent decades, e.g., by Hayami and Ruttan (1971) and Binswanger *et al.* (1978). Another important contribution to the autonomy debate was that of Jacob Schmookler (1966), who carried out a series of longitudinal studies (railroads, farm implements, petroleum refining, and paper-making) to elucidate the relationship between economic activity and "inventive activity", as measured by patents. Schmookler's work strongly supported the endogenous view. Recent work using patent data is also reported by Freeman (1982: pp. 53–70). Freeman, however, is somewhat critical of Schmookler's use of patent data, at least as a measure of "radical" inventions. A useful summary and critique of the literature can be found in Mowery and Rosenberg (1979), reprinted in Rosenberg (1982: Chapter 10).

[5] For instance, the continuous catalytic cracking process in petroleum refining was developed by a consortium organized for the explicit purpose of inventing around the Houdry process (Enos, 1962).

[6] The so-called experience curve was originally a purely empirical observation about trends in production costs as a function of cumulative production, or "experience". This idea was formulated as early as the mid-1930s, based on data from the aircraft industry (Wright, 1936). The relationship was applied to national economic performance by the Dutch economist Verdoorn (1949, 1951, and 1956), who formulated a "law" relating labor productivity and cumulative output. Arrow (1962) reformulated the notion in a much more comprehensive theory of economic progress, emphasizing the importance of learning in the capital goods industry and the

embodiment of technological progress in capital, resulting in increased labor productivity. There is a considerable literature on the question of whether technological progress can best be considered to be embodied in capital or labor, or whether it is more nearly disembodied. For instance, see Beckmann and Sato (1969). The results are essentially inconclusive. A progress function relating labor productivity and output, postulated by Kaldor (1961), was sometimes called "Kaldor's law". There is a sizeable empirical literature on the subject of progress functions or experience curves, e.g., Alchian (1963), Hirsch (1956), Rapping (1965), and David (1970). In recent years "experience curves" have been widely promoted as a strategic management tool, e.g., Cunningham (1980).

[7] Well-documented examples include hand-held calculators, digital watches, computer games, and personal computers. The number of competitors sometimes falls spectacularly at this point; see, e.g., Utterback (1986). Rosegger and Baird (1987) have documented the entry and exit of makes in the automobile industry.

[8] The diffusion literature is extremely voluminous. The first theoretical treatment in the economics literature was that of Mansfield (1961 and 1968), who primarily emphasized the demand side and the role of profitability for users. His original model assumes profitability to new users remains constant over time and, implicitly, that the "diffusion levels reached in later years also represent active adoption prospects during earlier years" (Gold, 1981). For a modern critique, see Metcalfe (1981). A number of attempts have been made to classify users into groups with different characteristics, such as "early adopters", "late adopters", and so on, e.g., Bass (1969). The "supply side" of inventive activity has been discussed more explicitly by Machlup (1962) and Nordhaus (1969), among others. Recent state-of-the-art reviews can be found in Mahajan and Wind (1986), from the marketing perspective, and Stoneman (1983), from the economics perspective. See also Wyatt (1986).

An independent literature has evolved around the problem of forecasting technological change *per se*, e.g., Jantsch (1967) and Ayres (1969), with an interesting subliterature on the special case of technological substitution, e.g., Fisher and Pry (1971). There is an obvious need to bring these two strands together.

[9] According to so-called standard (Heckscher–Ohlin–Samuelson) static equilibrium trade theory, a capital-rich country, such as the USA, should export capital-intensive products and import labor-intensive products. In reality, the USA exports labor-intensive "high technology" products and imports capital-intensive products of mature industries (Leontief, 1954).

[10] For the moment, I do not attempt any formal definition. The implicit question of comparative valuation raised here is addressed more explicitly later.

[11] Much has been written about the history of the telephone. See, for example, Goulden (1970).

[12] Obviously, there are many cases where the initial market is, in fact, a government. This was true for aircraft.

[13] It has been argued, e.g., by Schmookler (1966) and Freeman (1982), that the ICE was *not* a prerequisite, because steam and electric cars were still competing in the market as late as 1905 or so. However, the enormous expansion of the industry (its adolescent period) occurred after 1908, when automobiles became cheap and reliable enough for ordinary people to use for transportation. Electric cars were easy enough to use but had inadequate range (they still do); while steam cars were too hard to start, consumed too much fuel, and needed water every few miles. These technical disadvantages of steam were later overcome (in the Doble, produced in the 1920s), but not at a competitive price.

[14] Pig iron consists mainly of a solid solution of iron carbide Fe_3C in iron; it contains 4.5% carbon by weight.

[15] The distinguished historian Elting Morrison (1966) called the Bessemer process "almost the greatest invention" because of its great economic and social impact.

[16] In fact, relatively safe and effective gas lighting also became available about the same time and was not finally displaced for another three decades or more.

[17] To support this point further, Rosegger (1986) has pointed out that Bessemer was aware that his process would work even better using pure oxygen instead of air. But oxygen could not be produced economically in bulk until the Linde-Franke process was developed in the late 1920s and the BOF itself was delayed until 1952.

[18] Contrary to popular legend, Babbage's project was probably not technically infeasible. But it was undoubtedly underfunded, and Babbage compounded the problem by constantly tinkering with the design (Shurkin, 1984).

[19] The critical parts of a machine tool have to be made to tolerances an order of magnitude lower than the required tolerances of the parts to be made by it.

[20] This project began in 1953, using the IBM 701 computer, with a CRT-based random access memory of 2048 words.

[21] A consent agreement with the U.S. Department of Justice, under antitrust laws, forced AT&T to license the transistor technology to all comers for a very modest royalty. The Japanese were particularly quick to seize the opportunity.

[22] The first electronic computer (ENIAC) was unveiled in 1946. The transistor was announced in 1948. The first transistorized computers appeared c. 1956.

[23] Gerald Holton (1962) used the "mineral vein" metaphor primarily in the context of scientific research. In an earlier book of my own (Ayres, 1969), Chapter 5 introduced a similar idea. Wyatt (1986: p. 122), has likened researchers (or inventors) to "fisherman working a fishing ground". Thomas Hughes (1971) has used a slightly different idea, likening the advance of science and technology to an army advancing along a broad front. "Barriers" in Hughes's metaphor are like "reverse salients" or pockets of resistance holding up the front. When such a pocket is encountered, very intense efforts may be made to eliminate it.

[24] The equation itself goes back to Verhulst (1844).

[25] The progress of a new technology can be thought of as a special case of substituting "knowledge" for ignorance, or lack of knowledge.

[26] The phrase was used by J.A. Morton, Vice-President of Bell Laboratories, to characterize the limits of electronic technology as seen in the late 1950s. The limits were overcome, in this case, by the invention of monolithic integrated circuitry and its subsequent evolution to LSI, VLSI, and ULSI.

[27] See, for example, the good discussion by Freeman (1982: Chapter 7).

[28] This issue is discussed quantitatively in Ayres and Mori (1987).

[29] A good discussion of the role of market uncertainties is to be found in Freeman (1982).

References

Abernathy, W.J. (1978), *The Productivity Dilemma* (Johns Hopkins University Press, Baltimore, MD).

Alchian, A. (1963), Reliability of progress curves in airframe production, *Econometrica*, **31**, October: 679–693.

Alderfer, E. and Michl, H.E. (1942), *Economics of American Industry* (McGraw-Hill, New York, NY).

Arrow, K.J. (1962), Economic welfare and the allocation of resources for inventions, in R.R. Nelson [ed], *The Rate and Direction of Invention Activity* (Princeton University Press, Princeton, NJ).

Ayres, R.U. (1969), *Technological Forecasting and Long-Range Planning* (McGraw-Hill, New York, NY).

Ayres, R.U. (1984), *The Next Industrial Revolution: Reviving Industry Through Innovation* (Ballinger, Cambridge, MA).

Ayres, R.U. and Mori, S. (1987), Time preference and the life cycle: The logic of long-term high risk vs. short-term low risk investment, *European Journal of Operations Research* (in press).

Bass, F.M. (1969), A new product growth model for consumer durables, *Management Science*, **15**, January: 215–227.

Baumol, W.J. (1982), Contestable markets: An uprising in the theory of industry structure, *American Economic Review*, **72**, March: 1–15.

Beckmann, M.J. and Sato, R. (1969), Aggregate production functions and types of technical progress: A statistical approach, *American Economic Review*, March.

Binswanger, H.P., Ruttan, V.W., *et al.* (1978), *Induced Innovation: Technology, Institutions and Development* (Johns Hopkins University Press, Baltimore, MD).

Blackman, A.W. (1972), A mathematical model for trend forecasts, *Technological Forecasting & Social Change*, **3**(3).

Blackman, A.W. (1974), The market dynamics of technological substitutions, *Technological Forecasting & Social Change* **6**, February: 41–63.

Brozen, Y. (1951), Inventions, innovation and imitation, *American Economic Review, Papers and Proceedings*, 239–257.

Cunningham, J.A. (1980), Using the learning curve as a management tool, *IEEE Spectrum*, June.

David, P.A. (1970), Learning by doing and tariff protection: A reconsideration of the case of the ante-bellum United States cotton textile industry, *Journal of Economic History*, **30**: 521–601.

Diebold, J. (1952), *Automation: Advent of the Automatic Factory* (Van Nostrand, New York, NY).

Easingwood, C.J., Mahajan, V., and Muller, E. (1983), A nonsymmetric responding logistic model for technological substitution, *Technological Forecasting & Social Change*, **20**: 199–213.

Enos, J.L. (1962), *Petroleum Progress and Profits: A History of Process Innovation* (MIT Press, Cambridge, MA).

Fisher, J.C. and Pry, R.H. (1971), A simple substitution model of technological change, *Technological Forecasting & Social Change*, **3**(1).

Floyd, A. (1968) A methodology for trend forecasting of figures of merit, pp. 95–109 in J. Bright [ed], *Technological Forecasting for Industry and Gobernment* (Prentice-Hall, Englewood Cliffs, NJ).

Freeman, C. (1982), *The Economics of Industrial Innovation* (MIT Press, Cambridge, MA), 2nd edition.

Freeman, C., Clark, J., and Soete, L. (1982), *Unemployment and Technical Innovation: A Study of Long Waves and Economic Development* (Frances Pinter, London, UK).

Gold, B. (1981), Technological diffusion in industry: research needs and shortcomings, *Journal of Industrial Economics*, March: 247–269.

Gold, B., Pierce, W.S., and Rosegger, G. (1970), Diffusion of major innovations in US iron and steel manufacturing, *Journal of Industrial Economics*, July.

Goulden, J.C. (1970), *Monopoly* (Pocket Books, New York, NY).

Hayami, Y. and Ruttan, V. (1971), *Agricultural Development* (Johns Hopkins University Press, Baltimore, MD).

Heertje, A. (1983), Can we explain technical change? in D. Lamberton, S. MacDonald, and T. Mandeville [eds], *The Trouble with Technology* (St. Martins Press, New York, NY).

Hicks, J.R. (1932), *Theory of Wages* (Macmillan, London).

Hirsch, W. (1956), Firm progress ratios, *Econometrica*, April.

Holton, G. (1962), Scientific research and scholarship: Notes toward the design of proper scales, *Daedelus*.

Hughes, T.P. (1971), *Elmer Sperry: Inventor and Engineer* (Studies in the History of Technology, John Hopkins University Press, Baltimore, MD).

Jantsch, E. (1967), *Technological Forecasting in Perspective* (OECD, Paris, France).

Kaldor,N. (1966), *Causes of the Slow Rate of Growth of the United Kingdom* (Cambridge University Press, New York, NY, and London).

Kindleberger, C. (1953), *International Economics* (Richard D. Irwin, Homewood, IL).

Kindleberger, C. (1962), *Foreign Trade and the National Economy* (Yale University Press, New Haven, CT).

Kindleberger, C. (1979), An American economic climacteric, *Challenge*, **16**(6), January–February: 35–44.

Kleinknecht, A. (1986), *Innovation Patterns in Crisis and Prosperity: Schumpeter's Long Cycle Reconsidered* (Macmillan, London).

Kondratieff, N.D. (1926), Die Langen Wellen der Konjunktur, *Archiv für Sozialwissenschaft und Sozialpolitik*, **56**.

Kondratieff, N.D. (1928), Die Preisdynamic der Industriellen und Landwirtschaftlichen Waren, *Archiv für Sozialwissenschaft und Sozialpolitik* **60**.

Kuznets, S. (1930), *Secular Movements in Production and Prices – Their Nature and Bearing on Cyclical Fluctuations* (Houghton-Mifflin, Boston, MA).

Kuznets, S. (1940), Schumpeter's business cycles, *American Economic Review*, **30**, June: 152ff.

Kuznets, S. (1978), Technological innovations and economic growth, in P. Kelly and M. Kranzberg [eds], *Technological Innovation: A Critical Review of Current Knowledge* (San Francisco Press, San Francisco, CA).

Lenz, R. (1962), *Technological Forecasting*, ASD-TDR-62-414, June (Wright-Patterson AFB, Dayton, OH).

Leontief, W. (1954), Domestic production and foreign trade; The American capital position reexamined, *Economics International*, **7**, February: 9–38.

Levitt, T. (1965), Exploit the product life cycle, *Harvard Business Review*, November–December.

Lotka, A.J. (1956), *Elements of Mathematical Biology* (Dover Publications, New York, NY) [NOTE: Original title: *Elements of Physical Biology*, 1924].

Machlup, F. (1962), The supply of inventors and inventions, in R.R. Nelson [ed], *The Rate and Direction of Inventive Activity* (National Bureau for Economic Research and Princeton University Press, Princeton, NJ).

Mahajan, V. and Peterson, R.A. (1985), *Models For Innovation Diffusion*, (Sage Publications, Beverly Hills, CA).

Mahajan, V. and Wind, Y. (1986), *Innovation Diffusion Models of New Product Acceptance* (Ballinger, Cambridge, MA).

Mansfield, E. (1961), Technical change and the rate of imitation, *Econometrica*, **29**(4), October: 741–766.

Mansfield, E. (1968), *Industrial Research and Technological Innovation* (W.W. Norton, New York, NY).

Mansfield, E. (1983), Long waves and technological innovation, *American Economic Review*, **73**(2), May: 141–145.

Marchetti, C. (1977), Primary energy substitution models: On the interaction between energy and society, *Technological Forecasting & Social Change*, **10**(4): 345–356.

Mensch, G. (1975), *Das Technologische Patt: Innovation Überwinden die Depression* [Series: Stalemate in Technology: Innovations Overcome the Depression], (Umschau, Frankfurt, FRG) [NOTE: English translation: Ballinger Cambridge, MA, 1979].

Metcalfe, J.S. (1981), Impulse and diffusion in the study of technical change, *Futures*, **13**(5): 347–359.

Morrison, E.E. (1966), *Men, Machines and Modern Times* (MIT Press, Cambridge, MA).

Mowery, D. and Roseberg, N. (1979), The influence of market demand upon innovation: A critical review of some recent empirical studies, *Research Policy*, **8**, April: 103–153.

Nelson, R.R. (1962), Introduction to National Bureau of Economic Research, in R.R. Nelson [ed], *The Rate and Direction of Invention Activity* (Princeton University Press, Princeton, NJ).

Nelson, R.R. and Winter, S.G. (1982), *An Evolutionary Theory of Economic Change* (Harvard University Press, Cambridge, MA).

Nordhaus, W. D. (1969), *Invention, Growth and Welfare* (MIT Press, Cambridge, MA).

Ogburn, W.F. and Thomas, D. (1922), Are inventions inevitable? A note on social evolution, *Political Science Quarterly*, **37**, March: 83–98.

Olson, M. (1982), *The Rise and Decline of Nations: Economic Growth Stagflation and Social Rigidities* (Yale University Press, New Haven, CT).

Peschel, M. and Mende, W. (1986), *The Predator-Prey Model: Do We Live in a Volterra World?* (Akademi-Verlag, Berlin, FRG).

Price, D. de S. (1963), *Little Science, Big Science* (Columbia University Press, New York, NY).

Rapping, L. (1965), Learning and World War II production functions, *Review of Economics & Statistics*, February.

Rosenberg, N. (1982), *Inside the Black Box* (Cambridge University Press, Cambridge, UK).

Rosenberg, N. and Frischtak, C.R. (1984), Technological innovation and long waves, *Cambridge Journal of Economics*, **8**: 7–24.

Rosegger, G. (1986), Personal communication.

Rosegger, G. and Baird, R.N. (1986), Entry and exit of makes in the automobile industry, 1895–1960: an international comparison, *Omega: International Journal of Management Science* [in press].

Salter, W.E.G. (1960), *Productivity and Technical Change* (Cambridge University Press, New York, NY and London).

Schmookler, J. (1966), *Invention and Economic Growth* (Harvard University Press, Cambridge, MA).

Schumpeter, J.A. (1912), *Theorie der Wirtschaftlichen Entwicklunger* (Duncker & Humboldt, Leipzig, FRG).

Schumpeter, J.A. (1943), *Capitalism, Socialism and Democracy* (Harper & Row, New York, NY) [NOTE: Reprinted in Harper Colophon Edition, 1975].

Schumpeter, J.A. (1939), *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, 2 Volumes (McGraw-Hill, New York, NY).

Sharif, M.N. and Kabir, C. (1976), A generalized model for forecasting technological substitution, *Technological Forecasting & Social Change*, **8**: 353–364.

Shurkin, J. (1984), *Engines of the Mind* (W.W. Norton, New York, NY).

Skiadas, C. (1985), Two generalized rational models for forecasting innovation diffusion, *Technological Forecasting & Social Change*, **27**: 39–61.

Stoneman, P. (1983), *The Economic Analysis of Technological Change* (Oxford University Press, Oxford, UK).

Utterback, J. (1986), Unpublished mss.

van Gelderen, J. (1913), Springvloed: beschouwingen over industriele ontwikkeling en prijsbeweging, *De Nieuwe Tijd*, **18**(4,5,6) [NOTE: cited in Kleindienst].

Verdoorn, P.J. (1949), Fattoriche Regolanolo Sciluppo della Produttivita del Lavoro, *L'Industria*, **I**: 3–11 [NOTE: Translation by G. and A.P. Thirlwall].

Verdoorn, P.J. (1951), On an empirical law governing the productivity of labor, *Econometrica*, **19**: 209–210.

Verdoorn, P.J. (1956), Complementarity and long range projections, *Econometrica*, **24**: 429–450.

Vernon, R. (1966), International investment and international trade in the product cycle, *Quarterly Journal of Economics*, May: 290–307.

Wiener, N. (1948), *Cybernetics: Control and Communications in the Animal and the Machine* (John Wiley, New York, NY).

von Bertalanffy, L. (1957), Quantitative laws in metabolism and growth, *Quarterly Review of Biology*, **32**: 217–231.

Wright, T.P. (1936), Factors affecting the cost of airplanes, *Journal of Aeronautical Sciences* **3**: 122–128.

Wyatt, G. (1986), *The Economics of Invention: A Study of the Determinants of Inventive Activity* (Research Institute of the Finnish Economy, Series A10, UK) [NOTE: distributed by Harvester Press, Brighton, UK].