Chapter 2

Technology: Concepts and Definitions

Synopsis

The chapter provides an overview of diverse conceptualizations and terminologies that have been introduced to describe technology and how it evolves. First, technology is defined as consisting of both hardware and software (the knowledge required to produce and use technological hardware). Second, the essential feature of technology – its dynamic nature – is outlined. Technologies change all the time individually, and in their aggregate, typically in a sequence of replacements of older by newer technologies. Finally, the chapter emphasizes the multitude of linkages and cross-enhancing interdependencies between technologies giving rise to successive technology "clusters", which are the focus of the subsequent historical analysis chapters. The most essential terminology distinguishes between invention (discovery), innovation (first commercial application) and diffusion (widespread replication and growth) of technologies. As a simple conceptual model the technology life cycle is introduced. In this model, new technologies evolve from a highly uncertain embryonic stage with frequent rejection of proposed solutions. In the case of acceptance, technology diffusion follows and technologies continue to be improved, widen their possible applications, and interact with other existing technologies and infrastructures. Ultimately, improvement potentials become exhausted, negative externalities apparent, and diffusion eventually saturates, providing an opportunity window for the introduction of alternative solutions. Technology diffusion is at the core of the historical technological changes of importance for global (environmental) change. This is why the main emphasis in this book is on technology diffusion, which also provides the central metric to measure technological change. Less emphasis is placed on the complex microphenomenon of technology selection. The main generic characteristics of technological change are presented and some generalized patterns of technology diffusion are outlined. The chapter concludes with a discussion of sources and mechanisms, i.e., the "who's and how's" of technological change.

2.1. From Artifacts to Megamachines

What is technology?[1] In the narrowest sense, technology consists of manufactured objects like tools (axes, arrowheads, and their modern equivalents) and containers (pots, water reservoirs, buildings). Their purpose is either to enhance human capabilities (e.g., with a hammer you can apply a stronger force to an object) or to enable humans to perform tasks they could not perform otherwise (with a pot you can transport larger amounts of water; with your hands you cannot). Engineers call such objects "hardware". Anthropologists speak of "artifacts".

But technology does not end there. Artifacts have to be produced. They have to be invented, designed, and manufactured. This requires a larger system including hardware (such as machinery or a manufacturing plant), factor inputs (labor, energy, raw materials, capital), and finally "software" (know-how, human knowledge and skills). The latter, for which the French use the term *technique*, represents the disembodied nature of technology, its knowledge base. Thus, technology includes both *what* things are made and *how* things are made.

Finally, knowledge, or *technique*, is required not only for the production of artifacts, but also for their use. Knowledge is needed to drive a car or use a bank account. Knowledge is needed both at the level of the individual, in complex organizations, and at the level of society. A typewriter, without a user who knows how to type, let alone how to read, is simply a useless, heavy piece of equipment.

Technological hardware varies in size and complexity, as does the "software" required to produce and use hardware. The two are interrelated and require both tangible and intangible settings in the form of spatial structures and social organizations. Institutions, including governments, firms, and markets, and social norms and attitudes, are especially important in determining how systems for producing and using artifacts emerge and function. They determine how particular artifacts and combinations of artifacts originate, which ones are rejected or which ones become successful, and, if successful, how quickly they are incorporated in the economy and the society. The latter step is referred to as technology diffusion.

For Lewis Mumford (1966:11) the rise of civilization around 4000 B.C. is not the result "of mechanical innovations, but of a radically new type of

---

[1] From the Greek τϵχνϵ (*techne*, art, the practical capability to create something) and λογοσ (*logos*, word, human reason). Thus, τϵχνολογια (*technologia*) is the science and systematic treatment of (practical) arts. In a most general definition technology is a system of means to particular ends that employs both technical artifacts and (social) information (know-how).

social organization: ... Neither the wheeled wagon, the plow, the potter's wheel, nor the military chariot could of themselves have accomplished the transformations that took place in the great valleys of Egypt, Mesopotamia, and India, and eventually passed, in ripples and waves, to other parts of the planet". To describe the organization of human beings jointly with artifacts in an "archetypal machine composed of human parts", Mumford introduced the notion of a "mega-machine", with cities as a primary example.

Some may consider such semantics as philosophical overkill and irrelevant for a book on technology and global change. Others might find in them confirmation of a general uneasiness that technology is something large, opaque, and pervasive, which constrains rather than enhances our choices. Nevertheless it is important to present at the outset the broad continuum of conceptualizations of technology. It emphasizes that technology cannot be separated from the economic and social context out of which it evolves, and which is responsible for its production and its use. In turn, the social and economic context is shaped by the technologies that are produced and used. And through technology humans have acquired powerful capabilities to transform their natural environments locally, regionally, and, more recently, globally.

The circular nature of the feedback loops affecting technological development cannot be stressed too much. All the numerous technology studies of the 20th century share one conclusion: it is simply wrong to conceptualize technological evolution according to a simple linear model, no matter how appealing the simplification. Technological evolution is neither simple nor linear. Its four most important distinctive characteristics are instead that it is *uncertain, dynamic, systemic*, and *cumulative.*

Uncertainty is a basic fact of life, and technology is no exception. The first source of technological uncertainty derives from the fortunate fact that there always exists a variety of solutions to perform a particular task. It is always uncertain which might be "best", taking into account technical criteria, economic criteria, and social criteria. Uncertainty prevails at all stages of technological evolution, from initial design choices, through success or failure in the marketplace, to eventual environmental impacts and spin-off effects. The technological and management literature labels such uncertainty a "snake pit" problem. It is like trying to pick a particular snake out of a pit of hundreds that all look alike. Others use the biblical quote "many are called, but few are chosen". Technological uncertainty continues to be a notorious embarrassment in efforts to "forecast" technological change. But there is also nothing to be gained by a strategy of "waiting until the sky clears". It will not clear, uncertainty will persist, and the correct strategy is experimentation with technological variety. This may seem an "inefficient" strategy for progress. To the extent that it is, it is one of the many areas in which writers have drawn useful analogies between technology and biology.

Second, technology is dynamic; it keeps changing all the time. Change includes a continuous introduction of new varieties, or "species", and continuous subsequent improvements and modifications. The varying pace of these combined changes is a constant source of excitement (and overoptimism) on the one hand, and frustration (or pessimism) on the other. As a rule, material components of technology change much faster and more easily than either its nonmaterial components or society at large. The main factors governing technology dynamics are, first, the continuous replacement of capital stock as it ages and economies expand and, second and most important, new inventions.

Third, technological evolution is systemic. It cannot be treated as a discrete, isolated event that concerns only one artifact. A new technology needs not only to be invented and designed, but it needs to be produced. This requires a whole host of other technologies. And it requires infrastructures. A telephone needs a telephone network; a car needs both a road network and a gasoline distribution system, and each of these consists of whole "bundles" of individual technologies. This interdependence of technologies causes enormous difficulties in implementing large-scale changes. But it is also what causes technological changes to have such pervasive and extensive impacts once they are implemented. From historical research we know particular periods of economic development correspond with clusters of interrelated developments in artifacts, techniques, institutions, and forms of social organization. These mutually interdependent and cross-enhancing "sociotechnical systems of production and use" (Kline, 1985:2–4) cannot be analyzed in terms of single technologies, but must be considered in terms of the mutual interactions among all concurrent technological, institutional, and social change.

Fourth and finally, technological change is cumulative. Changes build on previous experience and knowledge. Only in rare cases is knowledge lost and not reproducible. A new artifact, like a new species, is seldom designed from "scratch". (The beginnings of the space program are a notable exception.) Hence, technological knowledge[2] and the stock of technologies in use grow continuously.

The following chapters emphasize the dynamic, systemic, and cumulative nature of technological change. In describing the history of technological

---

[2] One question is how much of the growth in information represents growth in *usable* knowledge? Rescher (1996) argues unconventionally that (usable) scientific knowledge only grows with the logarithm of the brute volume of scientific information.

change we discuss technological diffusion (i.e., technology's dynamic nature) largely in terms of technological "clusters" or "families", thus also highlighting technology's systemic and cumulative characteristics. We relate these to pervasive transformations in the economy, the spatial division of production, and also to environmental impacts. We have the benefit of hindsight, which conceals to a large extent the considerable uncertainties prevailing at the beginning of each technology cluster.

No claim to originality is made in adopting the notion of technology clusters as the organizing principle here. In 1934 Lewis Mumford characterized four phases of sociotechnical development according to dominant materials and energy sources used from preindustrial times to the 20th century (Mumford, 1934). Mumford's clusters set a useful historical stage, and we will build on them later as we extend the history of technology up through the last 200 years.

2.1.1. Terminology

The Austrian economist Joseph A. Schumpeter distinguished three important phases in technology development: invention, innovation, and diffusion.

Invention is the first demonstration of the principal, physical feasibility of a proposed new solution. An invention is usually related to some empirical or scientific discovery, frequently measured through patent applications and statistics. However, an invention by itself often offers no hints about possible applications despite the technological romanticism surrounding the inventor's human ingenuity. Even where applications are apparent, as in the recent frenzy surrounding the discovery of high temperature superconductivity, an invention by itself has no economic or social significance whatsoever.

Innovation is defined succinctly by Mensch (1979:123) as the point when a "newly discovered material or a newly developed technique is being put into regular production for the first time, or when an organized market for the new product is first created". A distinction is frequently made between process and product innovations. The former refers to new methods of production, for example, the Bessemer process of raw steel production. The latter refers to directly usable technological hardware, for instance, consumer products such as video recorders and compact disc players.

Numerous attempts have been made to discriminate between innovations that might be labeled "radical" or "basic" and others considered of lesser importance. But such distinctions are *ex post* rationalizations. At the moment of innovation proper it is nearly impossible to guess the ultimate or potential significance of an innovation (cf. Rosenberg, 1996). This inherent uncertainty (or inefficiency) is reflected in the fact that only a small percentage of innovations eventually "make it". The success rate is comparable to that of biological mutations. It is an essential feature of the evolutionary character of technological change, and we will return to it later when discussing technology selection.

Diffusion is the widespread replication of a technology and its assimilation in a socioeconomic setting. Diffusion is the final, and sometimes painful, test of whether an innovation can create a niche of its own or successfully supplant existing practices and artifacts. Technology assumes significance only through its application (innovation) and subsequent widespread replication (diffusion). Otherwise it remains either knowledge that is never applied, i.e., an invention without subsequent innovation, or an isolated technological curiosity, i.e., an innovation without subsequent diffusion.

One can elaborate on this basic framework of distinguishing between invention, innovation, and diffusion, by identifying additional intermediary steps and important feedbacks. Different methods of knowledge generation can be distinguished. For example, research efforts are classified into basic and applied research. Distinctions can also be made between research, development, and demonstration (RD&D). Distinctions can be made between radical and incremental innovations. The latter label is given to continuous improvements that extend applications, lower costs, and transfer new technologies into different sociocultural settings. Such continuous improvements are especially important as new technologies, like all innovations, are initially rather crude, deficient, and imperfect. Therefore considerable effort (research, development, marketing, etc.) is required to sustain pervasive diffusion.

Anyone who has driven a Model T Ford will appreciate that the artifact that we call a car today is markedly different from, and definitively easier to drive, than a similar artifact produced at the beginning of the century. Or compare the first brand of instant coffee to the hundreds of varieties that now cater to different tastes in such diverse places as Austria, Brazil, France, Saudi Arabia, and the USA.

In short, nothing could be more misleading than a simple linear model of knowledge and technology generation. To be successful, innovations must be continuously experimented with, and continuously modified and improved. Suppliers and users must work together; information from the marketing department must be fed back to the research lab in order to suggest new promising avenues for both applied and basic research. The appropriate metaphor or model is therefore that of *networks,* operating to generate innovations and to modify and tailor them in the course of diffusion.

2.1.2. Invention and innovation: Chronology and lags

Table 2.1 gives an abridged chronology of the development of railways, a particularly important technological innovation of the 19th century. The chronology is a good example of a long evolutionary line of developments with important precursor technologies and infrastructures. For example, the innovation represented by Stevenson's steam locomotive plant and the first 20 km Stockton & Darlington railway line in 1825 cannot be understood independent of earlier important developments in stationary steam engines and mine railways. *Table 2.1* also illustrates the considerable time lags that can take place in technological developments. For example, 55 years passed between invention and innovation dates of railways.

Although the timing of particular historical events is indeed important, most dimensions of technological development are continuous rather than discrete. They are either rooted in precursor technologies or rely on a confluence of various streams of developments, like the marriage of a new mobile power source (the steam locomotive) to an entirely new infrastructure system (rails). It is particularly the confluence, complementarity, and synergy between various streams of developments that characterize technological evolution. As a simple illustration consider a new product for which applications need to be found, production processes need to be established, materials must be chosen, and so forth. These activities require time and effort, and unless all aspects are addressed successfully, the new innovation may never appear on the market.

Table 2.2 shows a similar chronology for Neoprene, a synthetic rubber used, for example, in diving suits. In this case, more than two decades elapsed between invention and innovation. *Figure 2.1* indicates that, in general, decades are indeed the appropriate unit for measuring invention–innovation lags.

Figure 2.1 also reveals substantial variability. Of the 140 major innovations analyzed by Rosegger, 20 have lags over five decades, but nine have lags of less than a year. *Figure 2.1* includes innovations ranging from the electric railway, the jet aircraft, the telephone, and the transistor, to DDT, dynamite, margarine, and insulin. There is no clear decrease over time of the invention–innovation lags shown in *Figure 2.1.* Any advantage of modern organized R&D at the corporate level must therefore lie with other kinds of innovations rather than those traditionally considered in samples, such as that of *Figure 2.1*, of "basic" or "major" innovations. [Other examples are given in Mensch (1979:124–128) and van Duijn (1983:176–179). For a critical discussion, particularly of the Mensch sample, see Freeman *et al.* (1982) and Kleinknecht (1987).]

Table 2.1: A chronology of invention, innovation, and diffusion of railways.

Year	Event
1769	Watt patents low-pressure steam machine (invention)
1770	Cugnot develops steam-gun vehicle
1790	Read develops steam-powered road vehicle
1800	Watt's patent expires
1804	Evans constructs road steam locomotive
1813	Hadley develops locomotive to ride on rails
1814	Stephenson begins building locomotives
1820	About 40 private horse railways are operated between coal mines and the rivers Tyne and Wear in Northern England (Marshall, 1938)
1824	Stephenson builds first locomotive plant (innovation)
1825	Stephenson opens 20 km Stockton & Darlington line (beginning of diffusion)
1830	Opening of the Manchester–Liverpool railway, national railway network extends over 157 km
1845	UK railway network extends over 3,931 km; 0.2% of coal reaching London arrives by rail
1875	UK railway network extends over 23,365 km, transporting 490 million passengers and 200 million tons[a] of goods; 65% of London's coal arrives by rail
1900	UK railway network extends over 30,079 km
1900–1925	Railways achieve absolute dominance in UK transport market, transporting between 70% and 80% of all passenger- and ton-kilometers of the country; freight traffic reaches all-time peak with 570 million tons (including Ireland) in 1913; passenger traffic reaches its all-time high with 1.5 billion passengers in 1920
1928	UK railway network reaches maximum size with 32,846 km (end of diffusion and beginning of saturation and decline)

[a] Throughout this book ton is defined as metric ton, i.e., equal to 1,000 kg.
Source: Based on Marchetti (1980), and Grübler (1990a:90–122).

A few other illustrations of time lags include the example of nuclear energy in the USA; Fermi's Chicago reactor demonstrated the feasibility of a controlled nuclear fission reaction (invention) in 1942. It was not until 1957, 15 years later to the day after Fermi's demonstration, that the Shipping Port reactor went into operation (innovation).[3] It took over 30 additional years for nuclear reactors to account for 20% of US electricity generation. The prospects for further diffusion are highly uncertain.

---

[3] The sad military equivalent would be the first nuclear test bomb explosions and the first application in warfare, i.e., Hiroshima in 1945.

Table 2.2: Events in Neoprene development.

Year	Event
1906	Julius A. Nieuwland observed the acetylene reaction in alkali medium and worked for more than 10 years on the problem of higher yield of the reaction (invention)
1921	Nieuwland demonstrates that his material, "divinylacetylene", a polymer, can be produced through a catalytic reaction
1925	E.K. Bolton of Du Pont listens to a lecture of Nieuwland at the American Chemical Society; Du Pont assumes the further development of this type of rubber material
1932	E.I. Du Pont de Nemours and Company introduces Neoprene, a synthetic rubber, onto the market as a new, commercial product (innovation)

Source: Mensch (1979) based on Jewkes *et al.* (1969).

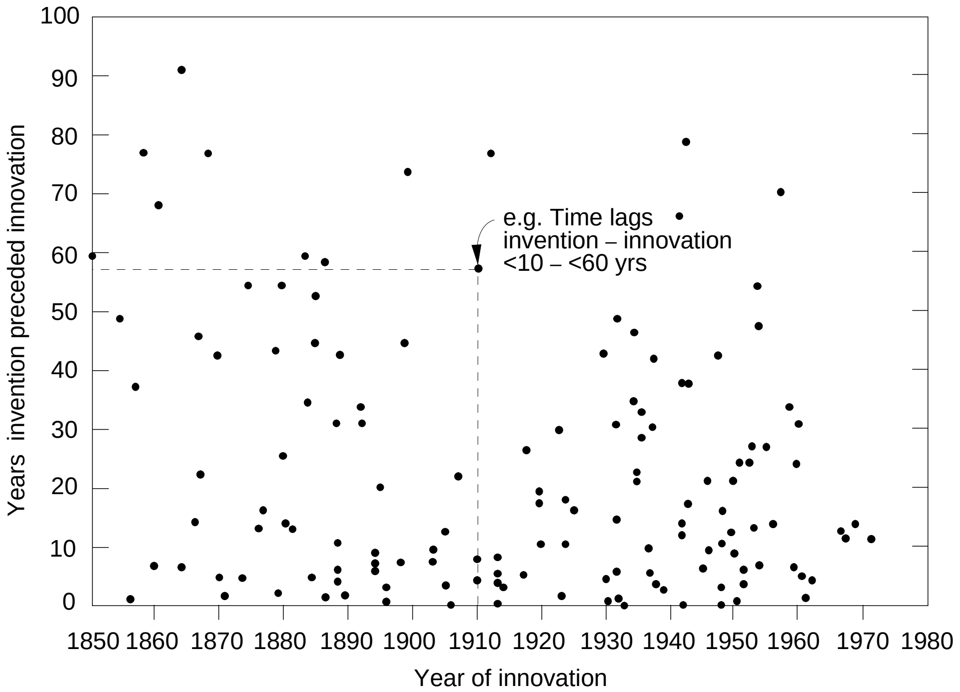


Figure 2.1: Time lag (in years) between invention and innovation of a sample of 140 major innovations introduced in the period 1850–1970. Source: Rosegger (1996:175).[1]

Postage stamps were first introduced in England in 1840 (innovation), but it took close to 50 years for a sample of 37 independent European, North American, and South American countries to follow suit (Pemberton, 1936). Compulsory school attendance in the USA was first introduced in 1847. It took until 1927 for the final state to follow suit.

These examples illustrate that changes in technologies and social techniques are not one-time discrete events. Technologies and techniques are neither developed nor changed instantaneously. Technology development is characterized by considerable time lags between development, first implementation, and widespread replication; all requiring considerable effort. Technology is not free. It is the result of deliberate research and development in university, government and private laboratories and by creative individuals. It requires cooperation between suppliers and users of new knowledge, between suppliers and users of technologies, and between proponents and opponents of particular technological solutions. Freeman (1994) provides an excellent review of recent research[4] identifying important linkages that exist between demand and supply, between users and providers of technology, between private and public R&D, and between knowledge and competencies internal to firms and those outside them. All of these shape the patterns and timing of invention and innovation.

2.1.3. The wider context of technology

In this section we present some general overall tendencies of technological evolution in the course of history. Counterexamples exist, and we admit that the discussion is not entirely free of our own analytical and personal biases. Nonetheless it provides a wider context of technological evolution that will be useful for the reader forming his/her own opinion of respective "progress"[5] or "regress" in the subsequent discussion.

Four general tendencies are identified:

- Increasing scale (cf. *Figure 2.2*), output, and productivity.
- Increasing variety and complexity.
- Increasing division of labor, both functionally and spatially.
- Increasing interdependence, interrelatedness and "network externalities".

These four tendencies should be seen not only as consequences of technological development, but also as resulting from technological "expectations"

---

[4] For a concise perspective from industry cf. Frosch (1984:56–81).

[5] For a critical appraisal of the value-laden concept of technical "progress", see Marx and Mazlish (1996).

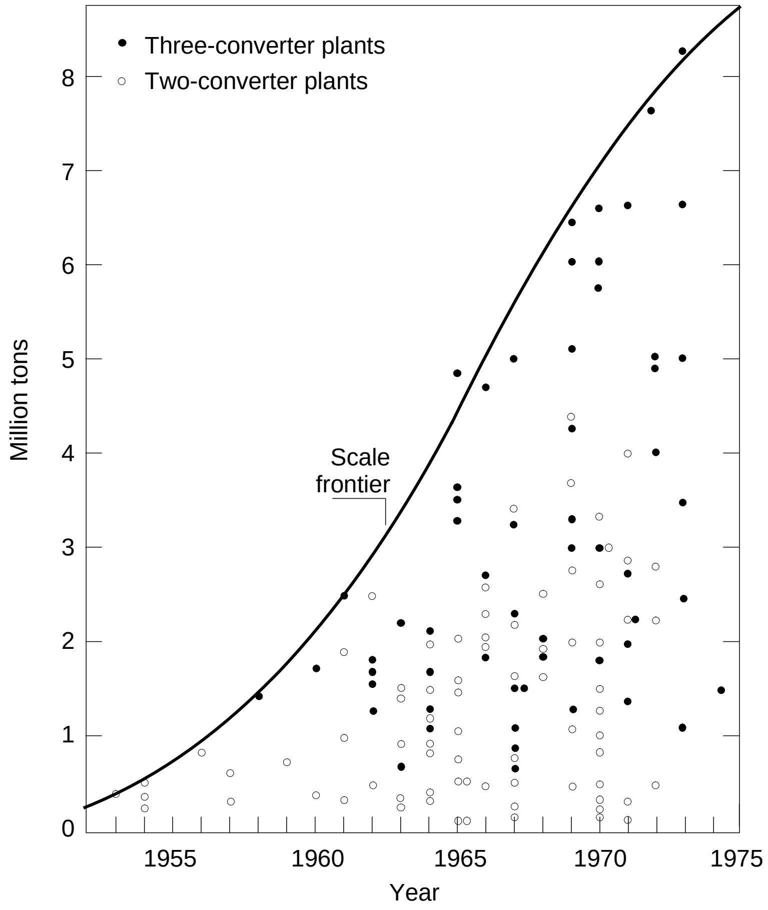


Figure 2.2: Increasing capacities of new steel plants (basic oxygen furnaces), in million tons. Source: adapted from Rosegger (1996:101).

(Rosenberg, 1982) that explicitly or implicitly shape the visions, missions, and expectations of those involved in the "technology business". We will return to this point in Section 2.3 when discussing entrepreneurship as a source of technological change.

Increasing Scale, Output, and Productivity

Increasing output, productivity, or efficiency is both a prime motivation and an effect when creating a new artifact. Increases can be quantitative or qualitative. A new production process can increase output either by scaling up existing production, or by reducing costs and thereby stimulating demand.

Economies of scale have been a pervasive phenomenon in increasing industrial output and lowering production costs. [Economies of scale exist when production costs increase less proportionally than the size of a production unit or a plant. Thus the costs of a 4-million-ton steel plant will be lower than the costs of two separate 2-million-ton plants. It is useful to distinguish (technology driven) economies of scale from (price driven) *economies of size.* In the latter, changing relative prices can lead to a different use of factor inputs, e.g., land versus labor in a farm, with technologies and production techniques otherwise unchanged. For instance, substitution of capital for labor in farming can lead to increasing farm size even in the absence of economies of scale proper.] *Figure 2.2* illustrates the extent to which the "scale frontier" has been pushed in oxygen steelmaking.

Other sources of output growth include growth in productivity and efficiency that enable to overcome resource limitations or to lower costs (and prices). Historically, growth in productivity and efficiency (lower input requirements per unit output) in most cases has led to increases in output rather than maintaining existing output levels and reducing inputs.

Improvements in economies of scale, productivity and efficiency do not come "automatically". They require engineering effort and experimentation. Such efforts and experimentation are an important source of technological learning and subsequent performance improvements.

A good example of an improvement that cut costs and stimulated demand comes from Henry Ford. With the assembly line he introduced standardized mass production to an industry characterized by small-scale production of customized items. That, after all, was how the automobile's predecessor, the horse carriage, had been produced. Reducing complex operations to a sequence of well-defined routinized jobs also enables better quality control and more focused learning and improvements in work routines. These, in turn, lead to further cost reductions.

Together with new materials (steel sheets), new forms of management and production organization (e.g., Taylorist time metering and optimization),[6] the Fordist assembly line reduced the selling price of a Model T Ford from US$850 in 1908 to US$290 in 1926 (Abernathy, 1978). This was possible despite increased wages to compensate for the increased work pressure that accompanied stepped-up output. The Model T production was standardized to such an extent that Henry Ford's quote that

---

[6] Frederick Winslow Taylor (1856–1915) developed a system of scientific management, primarily aimed at increasing labor productivity. The exact analysis and timing of production and work patterns, improvements in machinery, organizational changes, as well as financial incentives (bonuses) are characteristic elements of "Taylorism".

consumers “can have any color, provided it is black” became proverbial. Today even a “Fordist” assembly plant is run to provide substantial varieties of car models, colors, additional equipment, engines, and the like. New forms of production organization have also increased output, variety, and quality further. Volvo in Sweden, for example, pioneered a system combining assembly line operation with small assembly work teams. The result combines high output and productivity with more diverse and varied job responsibilities, thereby raising work satisfaction, lowering absenteeism, and raising productivity.

Output increases are not confined to industrial production. They also apply to new products and services. In industrialized countries, items such as the telephone, radio, television, home video recorder, and microwave oven are now standard equipment in most households. These expand people’s communication and entertainment options, both quantitatively and qualitatively. Enlarging consumer choices at reasonable costs creates precisely the demand to sustain increases in output. There is no mass production without mass consumption. Mass consumption, in turn, may have powerful environmental consequences – but that is a topic for a later chapter.

Finally, output increases qualitatively. Even if the number of cars or computers produced were constant, increases in performance, features, and designs would all increase output. Volumes and prices do not capture the full story of output growth. The comfort, safety, and reliability of today’s cars relative to their ancestors are as different as a Pentium PC from a 286 model, dubbed “advanced technology” at the moment of introduction. Both old and new “run”, but they “run” very differently. This presents serious problems in macroeconomic growth accounting, to which we will return when we turn to modeling issues. In emphasizing qualitative improvements we recognize it is not always easy to distinguish between quantity and quality. When consumers switched from black-and-white to color TV sets, for example, the black-and-white sets were often not scrapped. Instead they were moved to the basement or a secondary residence. Therefore, as a result of qualitative changes, the total number of TV sets in use increased also.

In addition to increasing output, technological change can also reduce inputs. Producing the same with less means a rise in productivity (efficiency), and historical productivity gains in terms of input reduction per unit of output have indeed been impressive. Industrial labor productivity (discussed in more detail in Part II) has increased by a factor of 200 or more since the middle of the 18th century. What took two weeks of work at 12 hours per day 200 years ago, is now produced in one hour. The energy requirements for producing a ton of iron or steel have dropped by a factor of more than 10 in the last 100 years.

Productivity gains are thus a central mechanism for improving the efficiency with which natural resources are used, and thereby reducing environmental impacts. But input reduction and output expansion often go hand in hand, and increases in productivity do not always lead "automatically" to resource conservation. Where productivity gains overcome resource constraints on further growth, output and its environmental impacts can expand. Technology is thus a double-edged sword in cutting the Malthusian resource limitation knot. Productivity increases have helped historically to overcome resource constraints so successfully as to expand output to unprecedented scales. Output has risen to such an extent as to face yet new limitations. Some are familiar input constraints on land, materials, and energy. But some are less familiar, such as limits on environmental capacity to absorb production and consumption wastes from ever larger output volumes.

Increasing Variety and Complexity

Another driver – and consequence – of technological change is increasing variety and complexity. Modern industrial systems produce not only a greater volume, but also an ever increasing variety of products. To the extent that variety multiplies a product's markets, it can generate cost reductions and profits. Thus economists speak of "economies of scope", in addition to economies of scale discussed previously.

The great variety of cars, computers, and travel packages to the remotest parts of the planet prove that mass production and standardization need not mean standardized products. There needs to be a functioning market that responds to consumer tastes for variety, as evidenced by the limited variety of consumer products in the former USSR. And much product variety may be classified as "pseudo-innovation", providing superficial variations in design or color, serving competitive and advertising strategies of firms. Consider the differences in the results of using alternative detergents in comparison with, for example, the marketing and advertisement effort devoted to different brands. Variety is exploding. The average number of items on sale in a typical large US supermarket has increased from 2,000 in 1950 to 18,000 items in the 1990s (Ausubel, 1990). The number of new items introduced into US grocery stores in 1993 alone totaled 17,000 (Wernick *et al.*, 1996). Of course, not all were successful. Westinghouse Electric Co. produces over 50,000 different steam turbine blade shapes, and the IBM Selectric typewriter, consisting of 2,700 parts, could be made in 55,000 different models (Ayres, 1988).

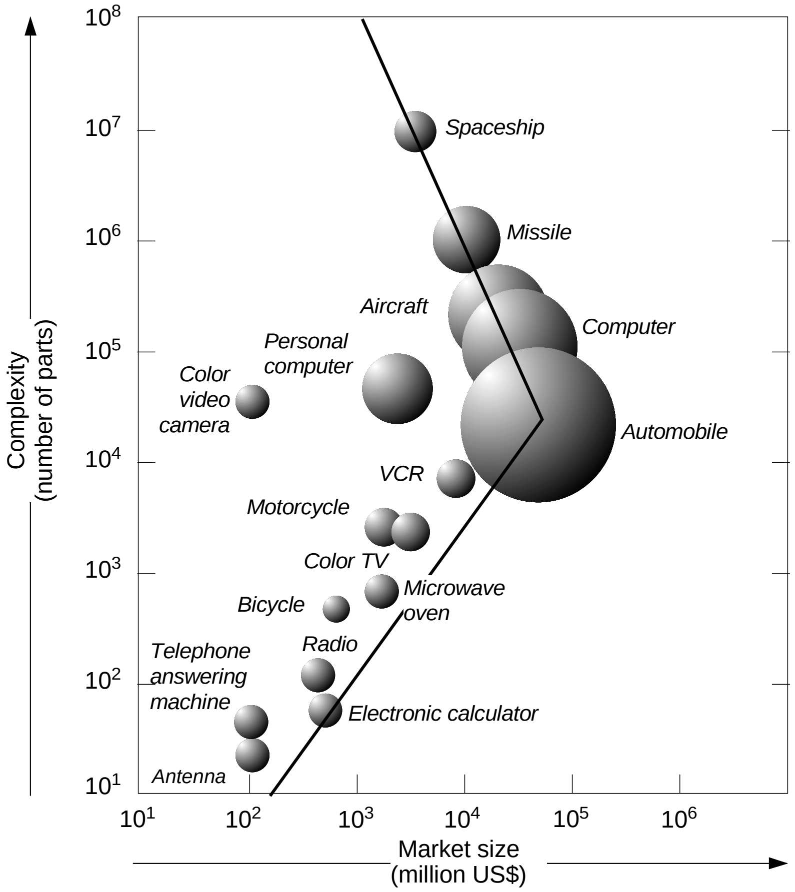


Figure 2.3: Market size (million dollars) and complexity (number of parts per item) of major durables produced in the USA. Source: Ayres (1988:28) based on Nagayama and Funk (1985).

Complexity is also increasing. Early hand tools like hammers, tongs, and shears typically involved two or three parts. A late 19th century hand drill accommodating various drill diameters involved 20 parts. A modern electric drill, including the motor, may have up to 100 parts. Vehicles are even more complex. The 1885 Rover safety bicycle consisted of approximately 500 parts, a modern car involves as many as 30,000 component parts, and a Boeing 747 roughly 3.5 million (all data from Ayres, 1988). The apogee (and nightmare) of mechanical complexity is the space shuttle with 10 million parts (see *Figure 2.3*).

Increased complexity means increased risk of production errors and consequent failures. A car with 50,000 components, and a failure rate of 1 per 1,000,[7] means 50 defective components per car. Inspection and quality control systems eliminate many defects, and design safety margins reduce the consequences of those that slip through. Ayres (1988:29) estimates that a single large US car manufacturer provides three billion opportunities for human assembly line error per day. Even with quality control and inspections reducing undetected errors to one in a million, the result would be 3,000 serious undetected production flaws per day, or about one in every three cars. Consumer surveys repeatedly report several manufacturing defects per car, although most are minor. Design safety margins, multiple inspections, and quality controls can be successful in reducing defects and their consequences. But eventually they are limited either by extreme complexity (as in the Challenger space shuttle), or in the case of an aircraft or nuclear power plant, by catastrophic consequences of failure (cf. Perrow, 1984). Multiple safety and backup systems are the usual response strategies, but they come at considerable additional cost.

Production risks due to complexity are only one part of the story. The other is risk due to human error when using the technology. Such risks are perhaps orders of magnitude larger than those from design and manufacturing defects, and they too increase with complexity. The history of large industrial accidents (e.g., Seveso and Chernobyl) reinforces this estimate. Technology, in the form of monitoring, automatic safety shutdown, and detailed safety procedures and protocols, can help reduce risks, but can never eliminate them entirely. Recent trends toward miniaturization (nanotechnology) and biotechnology promise reduced complexity. But biotechnology is still in its early stages, and may yet prove dauntingly complex. Living organisms like humans are, after all, several orders of magnitude more complex than even the most complex technological artifacts.

Increasing Division of Labor

Increasing complexity, sophistication, and skill requirements in both producing and using technologies require specialization. Metal tools, pottery and textiles have long been produced by specialists: craftsmen and craftswomen. Services have also long been provided by specialists: doctors, astronomers, accountants, writers, etc. In economics this specialization is called *division of labor*, enabled by increases in market size as described by Adam Smith in 1776 in his *Wealth of Nations.*

---

[7] The photocopier manufacturer Xerox heralded the success of a substantial reduction on its parts reject rate from 8 to 1.3 per 1,000 (Ayres, 1988:26).

Specialization and division of labor are pervasive phenomena of all societies beyond the neolithic period, so much so that numerous family names like Smith and Miller derive from an ancestor's trade. Historically, the trigger to specialization was a sufficiently large market size, and the spatial concentration of demand, specifically in the form of cities. In the industrial age, output growth and large-scale trade, via modern transport and communication systems, have much the same effect.

Since the transformation to an industrial society, the number of specialized professions has grown substantially. The yellow pages of any larger city, such as Vienna, contain more than 5,000 specialized trades, businesses, and services. Each subdivides into many further professional specializations.

Cities also provide the earliest examples of spatial division of labor. All those listings in the Vienna yellow pages presuppose the existence of a market where the supply of specialized job opportunities and the demand for specialized trades can meet. A book dealer, specializing in antique books of astronomy and geography, may find enough customers in a large city like Paris or New York, but certainly not in a village in the Tyrolean Alps. But spatial division of labor also results from differences in resource endowments and climatic conditions. Copper is mined where deposits are found, and tropical fruit cannot be grown in temperate climates. Much spatial division of labor results from economics. Production moves to where total costs are lowest. All costs need to be considered. An industrial plant can only be located where highly skilled labor is available. Transportation costs and the size of markets can be critical. In many specialized activities "intangible" factors such as proximity and close interaction with clients are important. This explains the existence of "high-tech" zones with high spatial concentrations of specialized firms in the computer and aerospace industries. Taken together, all these factors make location decisions highly complex and worthy of study by geographers, regional scientists, economists, and sociologists. Location decisions also entail a great deal of irreversibility because of the high sunk costs that result in terms of buildings, infrastructure, and personnel recruitment.

Spatial division of labor occurs at all levels: local, regional, national, and international. Many street names in European cities preserve the concentration of specialized trades that once resided there: goldsmiths, butchers, tailors, and traders. "Rustbelts" bear witness to the concentration of the coal, iron, and steel industries in regions of North America and Europe that "rusted away" when these industries declined. But perhaps increasing spatial division of labor is best illustrated by the increase in international trade (see *Table 2.3*).

Table 2.3: Index of growth in volume of world trade (1913=100).

ca. 1700	1
1800	2
1850	10
1900	57
1950	117
1970	520
1990	1,380
World trade (total exports f.o.b.) in 1990	US$3,397 billion
Distribution (%)	
Foodstuff	8.7
Raw materials	5.2
Energy	10.3
Chemicals	8.8
Machinery	35.7
Other manufactured goods	31.3

Abbreviation: f.o.b., free on board.
Source: Rostow (1978:669), Kennedy (1987:414), and IMF (1996:111). For a critical discussion of data sources of these historical estimates see Rostow (1978:663–669).

Total world trade in 1990 was around US$3,400 billion, or 13% of world GDP.[8] Trade is dominated by manufactured goods (75%, including chemicals) and by exports from industrialized countries (72%), mostly among themselves (57% of all world trade). Conversely, the share of primary resources including energy is less than 25% and the share of developing countries is also less than 25%. This asymmetry reflects the much smaller economic output in developing economies, plus low prices for raw materials relative to manufactured goods, thus the unfavorable "terms of trade" experienced by the developing world.

Increasing Interdependence and Interrelatedness

The final and fourth category of features that both drive technological evolution and are a consequence of such evolution covers technological interdependence and interrelatedness. Although difficult to describe and to model, the basic idea is that technologies increasingly depend on one another for both production and use. Consider the personal computer. It is built of hardware that needs to be produced and assembled. To run it, you need software. Switching it on requires an electricity network, with power plants, fuel supply infrastructures, primary energy extraction, and more. Network

---

[8] US$ in this book refers to constant 1990 money and prices, unless otherwise stated.

surfing requires more hardware (a modem), software, a telephone line, a local telephone network, and the internet itself. To ecologists the notion that "everything depends on everything else" might be familiar. However, to students of technology and policymakers, interdependence and interrelatedness create formidable challenges. It is impossible to manage change through attention to just a few "key" technologies.

In fact, because of technological interrelatedness, it may even be easier to manage change where few technologies and related infrastructures exist, such as in many developing countries. Consider, for instance, the example of cellular or satellite telephones that can be put in place everywhere, compared to a conventional telephone network system. This is the essence of the argument that latecomers to development may have genuine advantages too in terms that they can "leap-frog" (Goldemberg, 1991) older technology systems altogether. Conversely, countries "locked-in" to large existing technology systems face difficulties to move rapidly to newer systems. A historical example (England) and model for such entrenchment in old technology systems was first given by the economist Marvin Frankel in 1955 (Frankel, 1955).

As a contemporary example, consider the introduction of "zero-emission" vehicles, already mandated in California. They are not a technological novelty. Applicable inventions and innovations have existed since the turn of the century. Thus the difficulty lies not in producing electric cars, but in solving the chronic problem of power supply and storage. Without significant progress in batteries, for instance, the speed and range of electric cars is severely limited and costs are high. And a new infrastructure is also required for charging or exchanging discharged batteries.

Technologies depend increasingly on infrastructures of transport, energy, and communication. The service these provide is much larger than the usually modest costs charged to users. We notice them most, however, when we miss them most – when they fail. Thus infrastructures and related technologies are important examples of what economists call "network externalities". Consider your telephone: even with all costs paid, it would be useless if only you owned a phone. Rather, the utility of your phone increases with the number of participants in the telephone network and the more people and services you can access, e.g., to enquire about a flight departure, to order a pizza, or to chat with family and friends. Because costs are shared among all participants of the network, but each participant has the full benefits (utility) of being able to communicate throughout the network, the real value of the service remains "exogenous" to the price paid by an individual. This presents serious issues when new infrastructure networks need to be put in place. The high initial costs are incurred when benefits are still comparatively low;

if no one is prepared to incur the initial set-up costs, future benefits cannot arise. Distributive issues are also raised because those who incur the initial high costs are not the same people who reap the ultimate full benefits.

Thus like the air we breathe, for which we pay nothing, but without which we could not exist, infrastructures create important "externalities". These can be ignored in the microeconomic calculus, but they cannot be ignored by those studying or aspiring to direct technological change.

With the terminology and these four central tendencies of technological change in place, we can now turn to the most exciting feature of technology: technological dynamics or the mechanisms and patterns of technological change over time.

2.2. Technological Change

Some 10,000 years ago humans survived as nomadic hunters and gatherers. This required considerable sophisticated (technical) knowledge. (If you doubt this, try making a living today by hunting and gathering.) However, the first revolution in technology – the development of agriculture – changed the nomadic lifestyle dramatically. The development of markets and of money (institutional and organizational innovations or "technologies" in a larger sense) set people free from the need to be self-sufficient, enabling them to benefit from division of labor and specialization. Markets and agriculture (more precisely agricultural surplus production) were fundamental drivers for the emergence of cities.

Since that time, many further technological revolutions in fields such as materials, construction, navigation, and military technology have dramatically influenced the course of history. The past 300 years – the "age of technology" – have witnessed more momentous technological changes than any previous period in human history. Anthropologists, historians, and philosophers were quick to take an interest in technology and its role in shaping societies and cultures. Surprisingly, economists only came later to the study of technological change (Rosegger, 1996). Observing the Industrial Revolution from its midst, classical writers in economics from Adam Smith to Karl Marx could hardly fail to see the importance to economic growth of technological change, of new products and new production processes. But technological change – the "industrial arts" – was not seen as an integral element of the economic process. Even Karl Marx, who argued that transformations in the material structure of production determined changes in social relations, and who wrote extensively on technology, said relatively little about the sources of such changes (Rosegger, 1996).

Two economists deserve special credit for pioneering our thinking on technology: Thorstein Veblen and Joseph A. Schumpeter. Veblen (1904, 1921, 1953), perhaps best known for his *Theory of the Leisure Class* (first published in 1899), was the first to focus on the *interactions* between humans and their artifacts in an institutional context. He considered technology not as an exogenous force on entrepreneurs, engineers, or workers, but rather part of material and social relationships. Technology was developed and shaped by social actors, while at the same time shaping social values and behavior. Such a "circular" model of interactions was revolutionary at a time when technology was viewed as the exclusive domain of inventors, engineers, and "heroic" entrepreneurs (a kind of naive, romantic fascination adhered to even by the early Schumpeter). Such a unified view of technology contains a revolutionary message today, when many social scientists are trapped in a futile polarization between extreme positions of technology shaping society, or in turn society shaping technology.[9]

More widely acknowledged are the contributions of the Austrian economist Joseph A. Schumpeter (1883–1950),[10] who started his successful scientific career in Austria, passed through failed stages as an entrepreneur, served a short, unsuccessful interlude as Austrian finance minister, and completed his career at Harvard University. Schumpeter's *Theory of Economic Development*, published in 1911 and translated into English in 1934, is a landmark in considering the sources of technological change as endogenous to the economy. His later publications, in particular the monumental *Business Cycles* (1939) and the still eminently readable *Capitalism, Socialism and Democracy* (1942), deepened and extended the treatment of technology in his earlier work.

For Schumpeter the essence of technological change is "new combinations", particularly those that represent a discontinuity, i.e., new combinations that cannot be achieved by gradual modifications of existing artifacts, practices, and techniques. This Schumpeterian notion of technical change is referred to as "radical" technical (as opposed to incremental) change below.

> ... to produce other things or the same things by a different method, means to combine these materials and forces differently. In so far as the "new combination" may in time grow out of the old by continuous adjustment in small steps, there is certainly change, possibly growth, but neither a phenomenon nor development in our sense. In so far as this is not the case, and new combinations appear discontinuously, then the phenomenon characterizing development emerges. ... [the latter] ... is that kind of

---

[9] These extreme positions are referred to as "technological determinism" (e.g., Gille, 1978) versus the "social construction" of technology (e.g., Smith and Marx, 1994).

[10] For an excellent biography on the life and work of Schumpeter, see Swedberg (1991).

> change arising from *within* the system which so displaces its equilibrium point that the new one cannot be reached from the old one by infinitesimal steps. Add successively as many mail coaches as you please, you will never get a railroad thereby. [Joseph A. Schumpeter, *Theory of Economic Development*, 1934:64–66]

For Schumpeter the essence of technological change is "changes in techniques and productive organization", i.e., changes in technological hardware and software. As the above quote emphasizes, such changes are inherently "nonlinear". They entail both quantitative and qualitative characteristics that cannot be produced by simply adding linearly "more of the same" to existing technologies and practices.

Schumpeter also draws an important distinction between changes that emerge from an accumulation of small gradual changes (referred to as incremental improvements in the next section) and those that represent radical "new combinations". He gives five examples (1934:66), listed as follows:

1. The introduction of a new good or product, or of a new quality of a good or product.
2. The introduction of new methods of production, not tested yet by experience in the relevant branch of manufacturing. New production methods may be based on a new scientific discovery, or on a new way of handling a commodity commercially.
3. The opening of a new market, either one that did not exist before or one that has previously not been entered.
4. Obtaining (Schumpeter uses the rather inappropriate term "conquest of markets") new sources of raw materials or semimanufactured goods. The new source may already exist, or it may have been newly created.
5. New forms of organization, e.g., the establishment or the break-up of a monopoly.

It cannot be stressed enough that any technological change, whether incremental or radical, arises from *within* the economic system as a result of newly perceived opportunities, incentives, deliberate research and development efforts, experimentation, marketing efforts, and entrepreneurship. Technological change does not fall like "manna from heaven". Schumpeter also emphasizes the nonequilibrium nature of new combinations. Technological change is not simply "more of the same"; it radically changes the relations between economic inputs and outputs, and it changes the constraints under which these can evolve.

As we will see in the next section most macroeconomic models still largely ignore these two fundamental features of technological change, that is: (i) evolution from *within* (i.e., technological change should not be exogenous to the model); and (ii) the inherently dynamic and nonequilibrium nature of technological change, which static equilibrium models fail to capture. With this up-front pessimism about the treatment of technological change in much of economic modeling, let us return to Schumpeter's own words:

> ... Capitalism, is by nature a form or method of economic change and not only never is but never can be stationary. And this evolutionary character of the capitalistic process is not merely due to the fact that economic life goes on in a social and natural environment which changes and by its changes alters the data of economic action; this fact is important and these changes (wars, revolutions and so on) often condition industrial change, but they are never its prime movers. Nor is its evolutionary character due to a quasi automatic increase in population and capital or the vagaries of monetary systems of which exactly the same thing holds true.
>
> *The fundamental impulse that acts and keeps the capitalistic engine in motion comes from the new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates* [italics added].
>
> ... The history of the productive apparatus of a typical farm, from the beginnings of the rationalization of crop rotation, plowing and fattening to the mechanized thing of today – linking up with elevators and railroads – is a history of revolutions. So is the history of the productive apparatus of the iron and steel industry from the charcoal furnace to our own type of furnace, or the history of the apparatus of power production from the overshot water wheel to the modern power plant, or the history of transportation from the mail coach to the airplane. The opening of new markets, foreign or domestic, and the organizational development from the craft shop and factory to such concerns as US Steel illustrate the same process of industrial mutation – if I may use this biological term – that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism. [Joseph A. Schumpeter, *Capitalism, Socialism and Democracy*, 1942:82–83]

After setting the scene about the importance and essence of technological change, we can now introduce the finer conceptual and terminological detail in the following section, which presents a taxonomy of technological change.

2.2.1. A taxonomy of technological change[11]

Incremental Improvements

Occurring more or less continuously across all industry or service activities, incremental improvements resulting from scientific research and development, engineering, and learning effects improve the efficiency of all factors

---

[11] This section is based on Freeman and Perez (1988) and Freeman (1989).

of production. Although the combined effect of incremental improvements is extremely important, no single improvement by itself will have a dramatic effect. The accumulation of small incremental innovations in long-term overall productivity growth is extremely important, but the steps of individual improvements are difficult to document in detail. As a rule they can be documented through resulting aggregate productivity increases. Typical examples include reduced labor, materials, or energy requirements. The associated model is the "learning" or "experience" curve – with accumulated experience, humans learn to make things better, faster, and with fewer defects (see Section 2.3). Economists call this "learning by doing" (Arrow, 1962) and "learning by using".

The extent and rate of such learning effects vary according to the kind of learning involved. Most importantly they are not "autonomous". They should not be represented as an exogenous time-trend function, as is frequently the case in models trying to capture technological change. Learning depends on the actual accumulation of experience. Without "doing" there is no "learning".

Radical "New Combinations"

Radical "new combinations" are discrete and discontinuous events. In recent decades they have usually been the result of deliberate research and development efforts in industry, government labs, or universities. They may make quantum leaps in productivity possible and overcome resource limitations. Or they may enable the development of entirely new materials and products. Although they depart radically from existing engineering practice and technologies, they nevertheless often tie in with existing industrial structures. They therefore require no radical changes in overall industrial organization, although they do necessitate changes at the level of plants or even industrial sectors. The introduction of the Bessemer process, offering the possibility of low-cost, mass production of high-quality steel in the 19th century, the introduction of nylon, or the contraceptive pill both in the 20th century, are illustrative examples. Despite their importance for individual industrial sectors or submarkets, their aggregate economic impact remains comparatively small and localized, unless a whole cluster of radical "new combinations" is linked together to give rise to entirely new industries or services.

Changes in Technology Systems

Under this heading we refer to far-reaching changes in technology, affecting several branches of industry or occurring across several sectors of the economy. Such changes combine both radical and incremental innovations with organizational and managerial changes.

Technological change in one part of the economy triggers corresponding changes both upstream and downstream in related branches. A good example is the introduction of industrial electric motors (cf. Devine, 1982). Before their introduction, factories would have used a central steam engine with power distribution via transmission belts. Electric motors provided a new versatile decentralized source of motive power. They changed, first, the entire organization of the shop floor. Second, they required changes upstream in the production and distribution of electricity. Without such substantial changes in organization, both on the shop floor and in upstream electricity supply, the electric motor's impact on productivity would have remained localized and limited.

Devine (1982) estimates that the impact of the electric motor was multiplied by a factor of three through such organizational changes. The overall energy efficiency of a steam engine, coupled with mechanical power distribution, according to Devine's estimates is between 3% and 8%. If only the steam engine is replaced by self-generated electricity, the overall energy efficiency remains at 3–6%. However, combining utility-generated electricity and decentralized unit drives raises overall energy efficiency to 10–12%, or by a factor of three at the lower end of the range. These estimates report 1920s efficiencies. Current overall energy efficiencies for industrial drive systems are on the order of 25–28% (Nakićenović *et al.*, 1990), twice as large as 70 years ago.

Clusters and Families

Some changes in technology systems are so far-reaching that they impact upon the entire economy and nearly every aspect of daily life. Such changes involve whole clusters of radical and incremental improvements and may incorporate several new technology systems. The development of the automotive industry, for example, was contingent on developments in materials (high quality steel sheets), in the chemical industry (oil refining), in production and supply infrastructures (oil exploration, pipelines, and gasoline stations), in public infrastructures (roads), and a host of other technological and organizational innovations. The growth of the industry was based on a new way of organizing production, i.e., Fordist mass production combined with Taylorist scientific management principles. These yielded significant real-term cost reductions, making the car affordable to a wider social strata. This changed settlement patterns, consumption habits of the population, leisure activities, etc. And the automobile is just one among many consumer durables now considered standard in industrialized countries.

Clusters of interdependent radical innovations and technology systems give rise to whole families of hardware and software innovations with associated new institutional and organizational settings. Together they multiply the effects of each other on the economy and society. Thus their collective effect is more than the sum of their individual contributions. It would be impossible to calculate overall impacts even if detailed data on individual components were to exist. Qualitative descriptions are more appropriate. In the literature such clusters have been analyzed under the headings of "general natural trajectories" (Nelson and Winter, 1977) and "technoeconomic paradigms" (Freeman and Perez, 1988). Such clusters drive particular periods of economic growth, and will provide the central organizing concept for this book's analysis of technology and global change.

A Schumpeterian (1935, 1939) perspective on long-term economic growth and technological change sees overall development coming in spurts, driven by the diffusion of clusters of interrelated innovations and interlaced by periods of crisis and intensive structural change.[12] The existence of a succession of a number of such clusters over time does not mean that there is a quasi-linear development path, e.g., from textiles to basic metal industries to mass-produced consumer durables as alluded to in Rostow's (1960) stage theory of economic growth. Instead, such clusters are time-specific phenomena. The success of any one (in terms of economic growth) and the drawbacks (in terms of environmental impacts) cannot be repeated quasi-mechanistically at later periods in history or in different socioeconomic settings.

We adopt the concept of technology clusters and families to distinguish broadly between various historical periods characterized by different driving forces and patterns of technological change and their impacts. Our interest in global change issues together with technological interrelatedness and interdependence explains why we have adopted a taxonomy and perspective

---

[12] Such discontinuous paths of economic development have been corroborated by empirical studies ever since the seminal contributions of Nikolai Kondratiev (1926) and Joseph A. Schumpeter (1939). They received revived interest in the periods of economic crisis in the 1970s and 1980s (see e.g., van Duijn, 1983; Freeman, 1983; and Vasko, 1987). Beyond the empirical corroboration of important historical discontinuities, however, the interpretation and theoretical explanation of such long waves of economic and social development remains fragmented and open to further research. In particular, debate continues, first, on whether we are dealing with a recurring or cyclical phenomenon endogenous to the economy, and, second, on what causes the long waves that have been identified. For an excellent collection of classical, seminal papers of long wave theory including critical writings, see Freeman (1996).

with a deliberately large boundary. There are disadvantages to such an approach; we cannot dwell on the detail of individual artifacts and techniques. Instead, we must analyze them as systems and address their characteristics, and the scale and quality of their global change impacts, as a whole that is more than just the sum of its parts. In Chapter 4 we present briefly empirical evidence on the existence and timing of technology clusters, and identify appropriate indicator technologies that can be used as *pars pro toto* for their respective technology clusters and families. We focus on four major technology clusters since the beginning of the Industrial Revolution and identify a possible fifth cluster that in the next millennium could transform our entire technological and material base.

2.2.2. A taxonomy of global change: Impacts of technological change

With respect to (direct and indirect) global change impacts we group technological changes into four categories: (i) those that augment resources; (ii) those that diversify products and production; (iii) those that enlarge markets (output); and finally (iv) those that enhance productivity.

Technological Changes that Augment Resources

The tremendous historical expansion of industrial production has consumed enormous amount of natural resources in the form of raw materials and fuels. Technological changes that augment the resource base have therefore been essential. These include technologies that facilitate the discovery of new resource deposits and that improve the accessibility and recoverability of existing resources; technologies that represent new resource inputs altogether; and finally technologies that substitute for existing material and fuel inputs. Technologies that increase efficiency (i.e., enable to produce more with less inputs) can also be considered to augment resources, but we will discuss them separately under the general heading of productivity.

The onset of industrialization in 18th century England is usually associated with the emergence of coal as a major new industrial fuel. Although coal had been used in the brewing industry and to evaporate salt brines since the 13th century, its use remained limited because of restricted access to coal resources and limited applications. Coal was basically used in the same way as the fuelwood it was supposed to replace. Mining concentrated on comparatively shallow deposits, and coal could only be transported from mines located near riverways and the seashore. Hence the use of the term "sea coal" well into the 19th century. Two important technological innovations changed this situation. First was Abraham Darby's discovery of the coking process through which pig iron could be produced using coal instead of increasingly scarce and expensive charcoal. Second, the invention of stationary steam engines (Newcomen-Savary) allowed water to be pumped from greater depths than had been possible previously with mechanical pumps driven by horses. This increased physical access to deeper coal resources. These two technological innovations in turn paved the way for numerous subsequent innovations. The coking process eventually gave rise to an entirely new coal-based chemical industry that included city gas and synthetic versions of dyes like indigo. James Watt improved the thermal efficiency of the Newcomen stationary steam engine. It subsequently was used in mines not only for lifting water but also as a power source for mechanization, thus lowering mining costs and improving the economic accessibility of coal resources. Most importantly it became a mobile power source for railways. This further improved access to coal deposits and drastically lowered transport costs. With railway transport coal finally became just coal, and was no longer "sea coal".

Petroleum is another example of a new resource that both replaced other materials/fuels in existing uses and opened up new uses. Petroleum, in the form of kerosene, was initially used as a substitute illuminant for dwindling supplies of whale oil.[13] With advances in petroleum refining and the emergence of the internal combustion engine petroleum became a major transport fuel and petrochemical feedstock. That led to its use as a substitute for a variety of raw material inputs to industry (synthetic fibers, rubber, plastics, etc.). That the petroleum industry has grown to its current dominant position, despite recurrent fears of immediate resource exhaustion ever since the early 1920s, is a powerful illustration of the impact of technological change on augmenting resources through improved exploration, discovery, and access to increasingly remote and difficult environments.

Finally, entirely new resources have been made available through technological change. While copper and iron ores have been exploited since antiquity, it was only the introduction of aluminum that made bauxite a major resource for metal supplies. Similarly, nuclear technologies turned uranium into a new energy resource.

Technological Changes that Diversify Products and Production

This is the most familiar impact of technological change. Just compare the numbers and kinds of products and technological "gadgets" in nearly every

---

[13] For a concise account of how the industry drove whales nearly to extinction, see Ponting (1991:186–191).

household in the industrialized world today to the situation some 100 years ago. Electric lights, refrigerators, telephones, radio, TV, video, computers, automobiles, air travel, antibiotics, and vaccines were all either completely unknown or just curiosities with no social or economic relevance. Technological change has also opened up new production options. With steel, for example, production can now draw upon a variety of input materials (e.g., virgin iron ore or recycled steel scrap), energy sources, reductants, etc. to better match available inputs to production requirements, to increase product differentiation (e.g., speciality steels), and to increase quality.

Continuous change in product specifications makes it difficult to measure quality improvements outside "high tech" products such as aircraft or computers for which well-defined performance characteristics exist. Quality measurement problems are particularly relevant for consumer products. Therefore most analyses of technological change impacts on consumer product quality focus simply on falling real prices. A notable exception is a careful study by Payson (1994) analyzing a range of consumer products and their specifications from Sears Roebuck catalogues between 1928 and 1993. *Figure 2.4* reproduces his key findings for five different consumer products. (Note the semilogarithmic scale of *Figure 2.4*.)

Payson's analysis shows significant quality improvements even in consumer products with a low technology content such as sofas and shoes. Typically product quality improves at 2–3% per year. For higher technology products, such as gas ranges (ovens) and air conditioners, quality improvements range from 7% to 9% per year (Payson, 1994:119). These quality improvements are on top of price reductions (reflecting falling production costs) that have enabled mass diffusion of such products into nearly every household in industrialized countries. These quality improvements are generally not considered in macroeconomic statistics, which therefore tend to significantly underestimate the true impact of technological change [cf. also Nordhaus (1997) on this point and for an interesting case study on the costs of light].

Increased diversity as a result of technological change is continually counterbalanced by another tendency of technological change: standardization. Product and process innovations increase diversity, but the push to reduce costs increases standardization. The balance may well change in the near future in the age of new information technologies. These create the possibility of breaking the dominant paradigm of industrial mass production of standardized products. The sort of customized, one-of-a-kind products that are characteristic of preindustrial, handicraft production may reappear in industrial production. Current increasing product differentiation in aircraft, automobiles, and even textiles reinforces such a scenario.

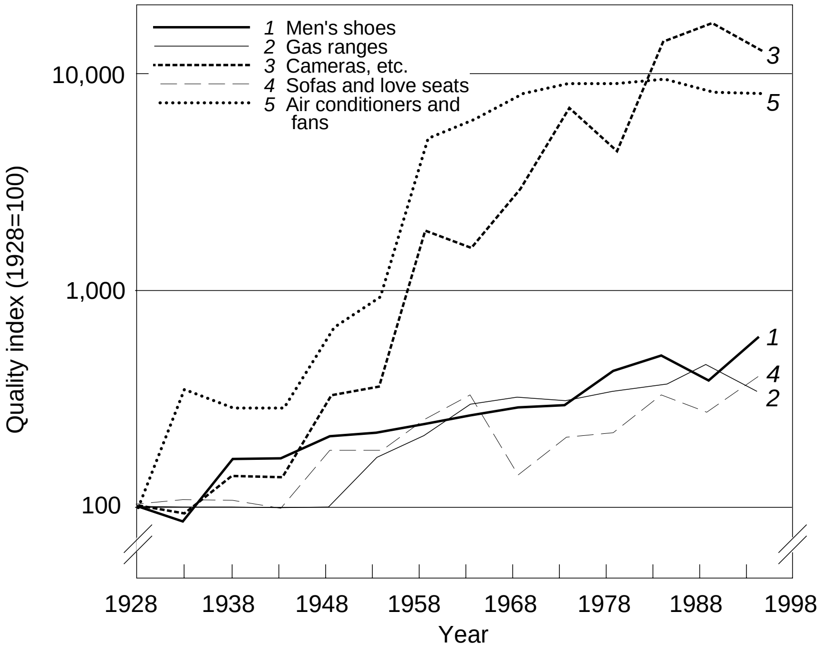


Figure 2.4: Evolution of the quality index (semilogarithmic scale) of five consumer products offered in the Sears Roebuck catalogues, 1928–1993. Source: Payson (1994:118).

Technological Changes that Enlarge Markets (Output)

Technological change has directly enlarged markets through successive transport revolutions from the canals, steam railways and ships of the 19th century to the road vehicles and aircraft of today. Higher transport speeds and falling costs have reduced the "economic" distance between production and raw material supplies on the one hand, and between production and markets on the other. These effects have enabled unprecedented increases in spatial division of labor through trade and market growth. Both permit increased economies of scale and have been important drivers in sustaining ever increasing output (and consumption) volumes.

Technological change also enlarges markets indirectly through improved productivity. Productivity improvements reduce production costs. Falling costs enable price reductions and expand the customer base and thus the market. The first automobiles and fax machines were expensive gadgets for a few wealthy individuals and institutions. With falling prices, the market for both products grew as they came within the financial reach of ordinary consumers. Mass consumption enables mass production, increasing economies of scale, further price reductions, and yet bigger markets. This positive feedback mechanism (here somewhat oversimplified) has driven the expansion of industrial production in domains as diverse as textiles, porcelain, cars, consumer durables, instant soups, electricity, and many more.

Technological Changes that Enhance Productivity

Productivity improvements are the key impact of technological change. Doing more with less is the central objective applying to all factors of production: land, labor, energy, and raw materials. Only with a long-term historical view can we grasp the scale of productivity increases due to continuous technological change over the last 200 years. The sources of these productivity increases are diverse and defy any simplifying summary. At this point, the key conclusion is simply that without such increases the spectacular historical expansion of human numbers, production and consumption could never have been sustained. It could not have been sustained in terms of resource availability, in terms of environmental impacts, or in terms of the economics of production and consumption.

In offering this simple taxonomy of technological changes we recognize the groupings are not clear cut. The impacts of technological change are frequently interdependent and overlap the categories defined above. We noted the relationship between productivity increases and expansions of the resource base and markets. It is similarly difficult to separate the direct impacts of productivity increases from their indirect impacts on mass consumption through increased wages and reduced working time. All are integral parts of the interwoven impacts of technological change that are relevant for global change, even if the impacts are too frequently subsumed under output growth and increasing environmental burdens.

2.2.3. Technological dynamics and interaction

The fact that the essential feature of technology is *change* causes an epistemological problem. In trying to describe a particular technology such as the railway or car, we have to face the problem that the object of our investigation keeps changing. Initially a new technology is imperfect, expensive, and limited in its applications. It must first prove itself in niche market applications where performance rather than cost is the overriding criterion. If successful, subsequent improvements and cost reductions can lead to wider applications. This evolution is the essence of the technology life cycle model described below. It is important to remember that the technology being analyzed in any particular case is only defined with the benefit of hindsight.

It is almost impossible to anticipate a new product's future applications or the new "combinations" that may become part of its life cycle.

To date no comprehensive method has been developed to describe and classify the myriads of technological artifacts and techniques. At the sectoral level, attempts have been made (e.g., Foray and Grübler, 1990) to use morphological analysis techniques, first, to describe the total evolutionary space of possible combinations capable of performing a specific task, and, second, to map the historical "branching" of the evolutionary tree of actual combinations. Such an analysis illuminates the functions that particular technological "combinations" can provide, and which combinations remain "locked out". It thus helps identify feasible, unexplored alternatives that may emerge later as possible "surprises" and competitors. However, such analyses are extremely data-intensive and therefore remain localized and very specific.

It is somewhat easier to classify technology dynamics than it is to classify technologies. As a first step, we simply consider the evolution of a particular artifact or technique with an "introspective" perspective, e.g., looking at its design features, performance, price, scale, and various productivity measures. This is the principal perspective of technology life cycle models. Second, we consider how a particular technology interacts with its environment: what are the factors determining its growth or failure; how does it perform in a particular market; and how does it complement or compete with other artifacts and techniques? This is the perspective of technology diffusion and substitution models. It is only through diffusion that inventive and innovative potentials are translated into actual changes in social practice, artifacts, and infrastructures. Diffusion phenomena are therefore at the heart of all changes in society and its material structures.

Technology Life Cycles

The world of technology is full of biological metaphors: for example, evolution, mutation, selection, and growth. Some are more appropriate than others. The clearest metaphor is between biological and technological growth or life cycles, and it is one that is widely used in the technological, management, and marketing literature.[14] The appeal of the life cycle model lies primarily in its considerable success as, first, an empirically descriptive tool and, second, as a heuristic device capturing the essential changing nature of technologies, products, markets, and industries. The essence of the technology life cycle model (like that of other growth models in biology) is that growth is nonlinear, and especially not unlimited. Typically growth in biology and

---

[14] For an excellent (and also critical) survey, see Ayres (1987).

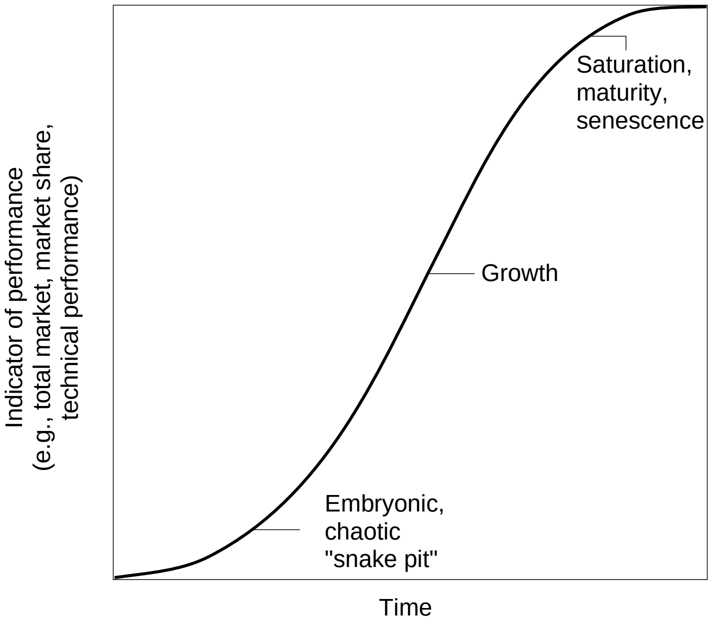


Figure 2.5: A stylized technology life cycle model.

of technologies alike proceeds along an S-shaped pattern: slow growth at the beginning, followed by accelerating growth that ultimately slows down leading to saturation. However, the S-curve or life cycle model is not an explanatory one. It does not explain *why* things evolve as they do.

The technology life cycle model (see *Figure 2.5*) classifies the phases of technology development into three phases: childhood, adolescence, and maturity. Subsequently, decline or senescence (and ultimate death) may follow. These correspond to a technology's introduction, growth, saturation, and eventual decline. Typically a technology's life cycle is described by indicators such as output volumes, market share, product characteristics (performance), sources of technological change, and the structure of industry. Most important with the last three of these is whether a life cycle phase is characterized by diversity or standardization. Associated with each of the three phases of the life cycle is a "stylized" pattern[15] as described below.

Introduction/childhood. The first phase is characterized by low production volumes and market shares and is the period with the greatest technological

---

[15] These patterns are "stylized" in that they represent a simplified summary of a large number of product and industry studies. In many individual cases deviations from these "stylized" patterns can occur.

and market diversity. Many possible technological designs are explored, development focuses on product innovations, and numerous firms try to gain a footing in the market. Emphasis is on demonstration of technical viability, and costs are of secondary importance. Learning effects and technology improvements derive primarily from experimentation and R&D. Overall the market is highly volatile and uncertain, characterized by a large number of "drop-outs", both of design alternatives and firms.

Growth/adolescence. Initial diversity gives way to increasing standardization as technical viability is established and efforts begin to be made to improve production economics. Increasing certainty of technological viability and applicability, reduced risks to innovators, and falling costs and prices lead to rapid market growth. Product innovations improve a technology's design features and enlarge its field of application. Process innovations improve production economics, and significant learning effects for both producers and users additionally reduce costs. Such innovations and learning effects provide positive feedbacks that further stimulate market growth. Eventually, however, the competitive environment becomes increasingly concentrated. This concentration applies first of all to firms and industry structure. Either because smaller firms go broke, or are absorbed in mergers and acquisitions, the number of producers declines rapidly. The history of the automobile industry is a case in point (*Figure 2.6*), although hardly an extreme example. For instance, there are fewer than five large commercial aircraft and aircraft engine manufacturers worldwide. Of course, *product variety* continues to be large, and is even increasing, as ever more specialized applications are searched (and found) for technologies and products.

Although the number of radically different designs diminishes in favor of a few demonstrated alternatives, these continue to be modified and adapted for increasingly diverse and remote applications. Whereas design changes in the early phases are characterized by a rapid succession of new models with increasing performance and productivity, later phases are characterized by incremental design changes. The passenger aircraft industry is a good example. Aircraft productivity, in terms of passenger-kilometers per hour, increased between the 1930s and 1970s through a rapid succession of different designs from the classic DC-3 of the 1930s to the Boeing 747 "jumbo" jet of the 1970s (*Figure 2.7*).

These rapid design changes allowed improvements to be made not only in aircraft productivity but also in fuel economy and crew productivity. Since 1970, however, improvements have been incremental. The B-747 has been "stretched" by increasing its length, stretching the double deck, and so forth. Incremental improvements can be impressive; a modern B-747 (400 series)

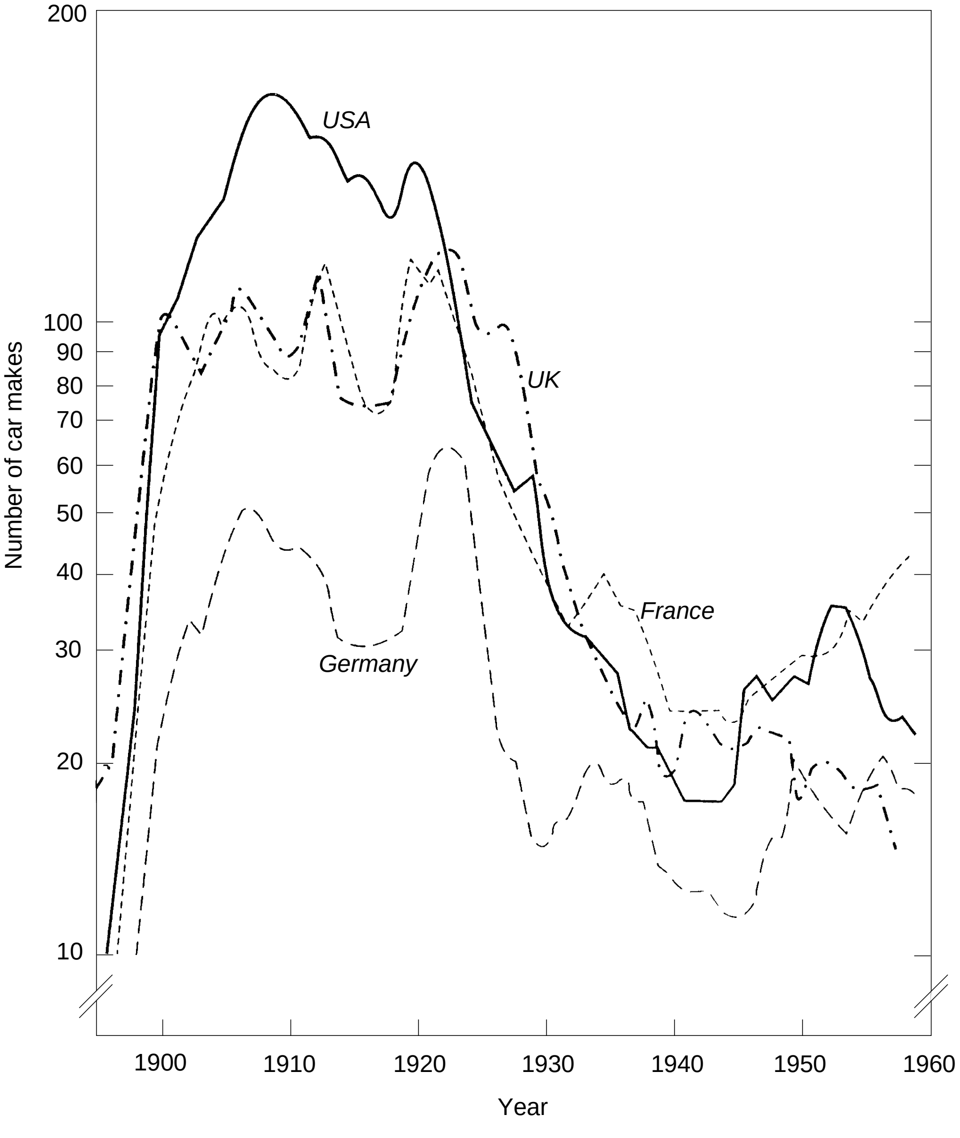


Figure 2.6: Number of car makes, 1895–1960 (on semilogarithmic scale), showing the increasing market concentration characteristic of a maturing industry. Note persistent differences between countries even under a similar overall trend of substantial reductions in car makes competing on the market. Source: Rosegger and Baird (1987:96).[2]

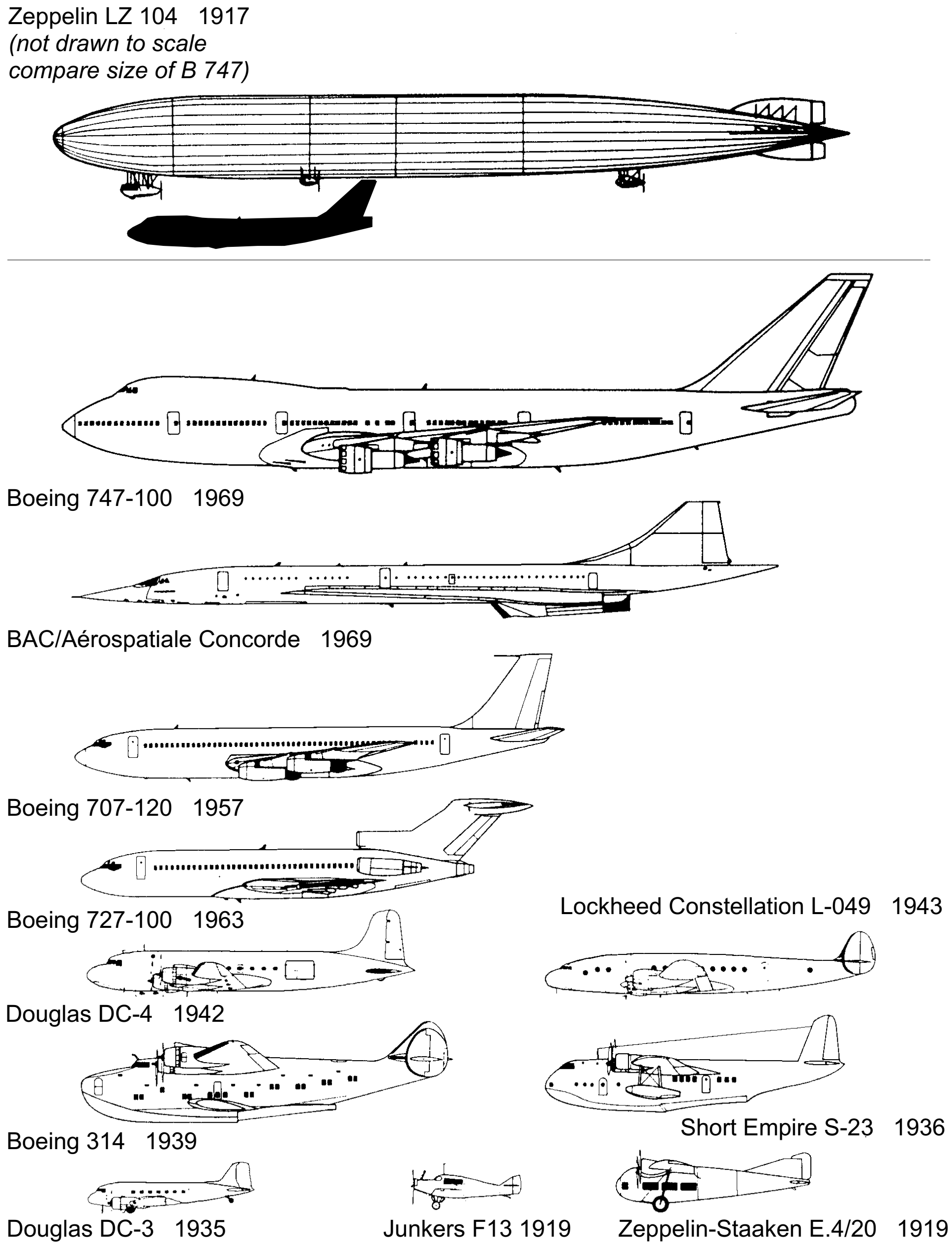


Figure 2.7: Size of selected commercial passenger aircraft. Note the comparatively modest size of today's commercially successful Boeing 747 jumbo jet relative to that of the unsuccessful Zeppelin from the beginning of the century. Source: Hugill (1993:256).[3]

consumes one-quarter less fuel than its 100 series counterpart of 1969 (Borderon, 1990:33). But the incremental nature of improvements reflects the increasing maturity of current aircraft technology, even if subsystems may continue to change radically (e.g., the new fly-by-wire system introduced in the Airbus 320/340 series).

Saturation/maturity. Growth rates slow down as markets become saturated and improvements face diminishing returns. Competition is based almost entirely on cost reduction rather than design improvement, and the market is concentrated in the hands of a few suppliers. The labor and skill intensity of production becomes increasingly "internalized" in machinery and mechanization. Large plants operate with almost no labor.

The management literature is full of examples of industries "taken by surprise" by market saturation and the slow down of market growth (e.g., Porter, 1983, 1990). Marketing departments typically continue to forecast a recovery in growth "just around the corner", and there are considerable lags in adjusting investment and expansion plans. As a result, the industry faces considerable overcapacity and intensified competition and market volatility. Common responses are to concentrate production to squeeze out the last marginal cost improvements from scale economies, or to outsource production altogether. This is one of the core areas of current concerns about job losses due to "globalization", but it should be related to increasing market saturation and industry maturity phenomena, rather than globalization per se. On the product side, design innovations focus on packaging and appearance rather than intrinsic features and qualities. The technology or product finally turns into a mass-produced commodity increasingly subject to regulation and an increasing awareness of its disbenefits. Disbenefits, such as environmental impacts, are generally either not anticipated in the earlier phases of a technology's life cycle or considered of secondary importance. Many problems also emerge nonlinearly with increasing application densities, and these in particular constitute genuine "surprises" (Brooks, 1986) to industry, consumers, and governments. The classic example is the automobile, which increases congestion and pollution as the number of them on the road grows. Thus, even small additional growth can suddenly generate important "externalities" that limit the usefulness of further growth.

We next turn to the mechanics of diffusion that underlie the progression through the three life cycle stages. As an initial illustration let us turn back the clock nearly 1,000 years and return to monastic life in 11th-century Burgundy.

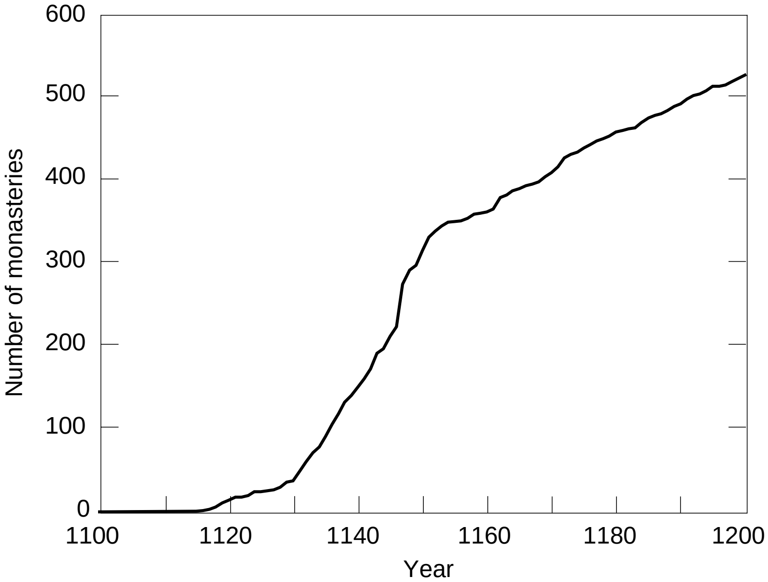


Figure 2.8: Diffusion of Cistercian monasteries in Europe: the first 100 years. Data source: Janauschek (1877).

A Medieval Prelude

In 1098 movement for the reform of Benedictine rule led St. Robert to found the abbey of Citeaux (Cistercium). Citeaux would become the mother house of some 740 Cistercian monasteries, about 80% of which were founded in the first 100 years of the Cistercian movement. Nearly half were founded between 1125 and 1155, and many traced their roots to the Clairvaux abbey founded as an offshoot of Citeaux in 1115 by the tireless St. Bernhard. The nonlinear, S-shaped time path of the spread of Cistercian rule (*Figure 2.8*) resembles the diffusion patterns we will observe later for technologies. In terms of the terminology introduced previously (Section 2.1.1), we might say that St. Robert invented Cistercian rule, St. Bernhard innovated, and diffusion followed. This basic pattern of temporal diffusion is essentially invariant across centuries, cultures, and artifacts: slow growth at the beginning, followed by accelerating and then decelerating growth, culminating in saturation. Sometimes a symmetrical decline follows.

Diffusion is a spatial as well as a temporal phenomenon. The topology of the Cistercian network reveals a hierarchy of centers of creation and structured channels of spread. *Figure 2.9* illustrates some example pathways in the spatial spread of two Cistercian "subfamilies", named after their respective mother houses as lines of Clairvaux and of Morimond.

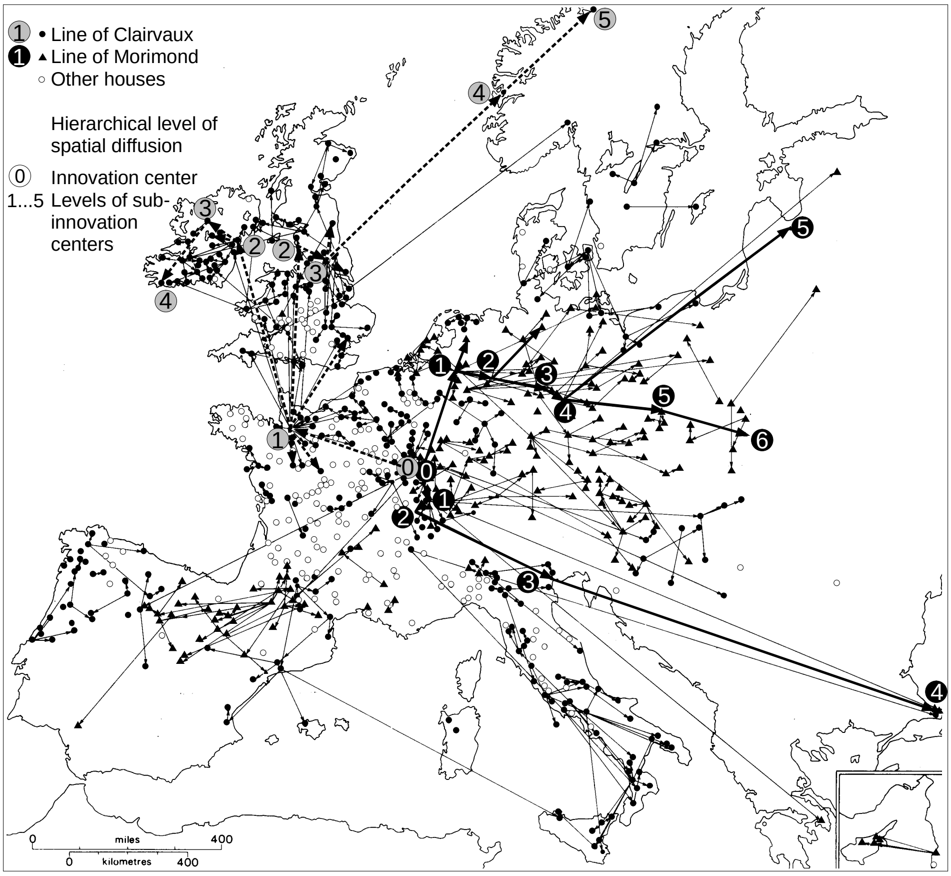


Figure 2.9: Spatial diffusion of Cistercian settlements (lines of Clairvaux and Morimond). Note in particular the hierarchical topology of spatial diffusion, from innovation centers to subcenters, and from the respective centers out to the hinterlands as illustrated for selected traits of the two houses. Adoption densities (settlements in this case) are highest in the innovation centers, and lowest in the hinterlands. Bottom right hand box shows diffusion to Cyprus. Source: adapted from Donkin (1978:28–29).

The patterns bear witness to the existence of networks, and today there is a growing literature on the role social and spatial networks play in the diffusion process (cf. Kamann and Nijkamp, 1991). *Figure 2.9* also shows significant differences in the spatial density of settlements. The origin of the innovation, Burgundy, was home to all four mother houses and had the highest spatial concentration of settlements. From there daughter houses were founded (regional "subinnovation centers" in the terminology of spatial diffusion), from whence Cistercians further spread to their respective hinterlands (the "neighborhood effect" in spatial diffusion) to found other subregional centers, which in turn led to further settlements.[16] The density of settlements decreases the further one moves away from the original center and from each subsequent regional and subregional center. The result is persistent regional differences and disparities.

The importance both of social networks and of diversity is exemplified by differentiation into different Cistercian "subfamilies". Each was named after its respective mother house, and each followed its own pattern of settlements, regional specialization, and implementation of Cistercian rule. Some additions to Cistercian rule were not genuine new settlements, but were "takeovers". For example, Savigny, with all its daughter houses, submitted to Clairvaux rule in 1147 and subsequently became the mother house of all Cistercian settlements on the British Isles. Despite differentiation and regional specialization, close communication existed between all the monasteries, creating an important channel for the spread of 13th- and 14th-century innovations like the water mill, new agricultural practices, and Gothic cathedral architecture.

The Cistercian movement had significant social, economic, and environmental impacts. It was particularly instrumental for the introduction of new agricultural practices and manufacture of textiles. Moreover, Cistercian rule commanded location of settlements in remote areas. This made Cistercian monasteries important local nodes for the internal colonization of land in Europe, and for early deforestation as well (see Part II).

Technological Diffusion and Substitution

Technological growth is the central feature of the technology life cycle, and is measured either in terms of growing volumes (e.g., tons of steel, number of cars) or growing market shares. Such growth cannot be analyzed by focusing narrowly on an artifact or product itself, but can be understood only by examining how a technology interacts with its environment, including other technologies. This interaction is the essence of technological diffusion and substitution. As illustrated in our medieval prelude, diffusion phenomena are not linked to the spread and growth of technological artifacts alone, but are

---

[16] Spatial diffusion proceeds in a kind of patchwork and hierarchical manner. Originating from innovation centers diffusion proceeds first to the areas in close proximity to the center (the center's neighborhood, or its "hinterland"). At the same time, the innovation is "exported" to other, more remote places (regional subinnovation centers) and spreads from there to the respective hinterlands as well as to further remote (third or even higher level hierarchical) subcenters of innovation diffusion. The classical work of spatial innovation diffusion remains the seminal book of Torsten Hägerstrand (1967).

a much wider social phenomenon (see Rogers, 1962, 1983). The most general definition of diffusion is: an innovation (idea, practice, artifact) spreads via different communication channels in time and space, among members of a social system. A primer on diffusion, as well as some elementary mathematics describing diffusion and growth, is given in *Box 2.1*.

Some instances exist of what might be called "pure" diffusion where an idea, practice, or artifact represents such a radical departure from existing solutions that it creates its own niche for diffusion. More frequently, however, a new solution does not evolve in a vacuum but interacts with existing practices and technologies. This is referred to as technological substitution,[17] with the new solution either competing one-on-one with an existing alternative or competing with several different technologies simultaneously. These interactions are usually best understood by examining relative (i.e., market) shares of competing alternatives, rather than absolute volumes.

Figure 2.10 illustrates the growth of the US canal network in the 19th century, along with other important transport infrastructures. The empirical data are approximated by a symmetrical growth curve (a three parameter logistic in this case).[18] The estimated asymptote (saturation or maturity level) of the diffusion processes of canals is approximately 4,000 miles and in good agreement with the actual maximum of 4,053 miles (6,400 and 6,485 km, respectively) reached in 1851 (shown as 100% diffusion level in *Figure 2.10*). The standard measure of diffusion speed is the time a process takes to grow from 10% to 90% of its ultimate saturation level (see *Box 2.1*). In the case of symmetrical growth this also equals the time required to grow from 1% to 50% of the saturation level.

In *Figure 2.10* the diffusion rate for canals, Δt, equals 31 years, and the entire diffusion cycle spans about 60 years. Thus, it took more than half a century to develop the canal network in the USA, with most canals (80%) constructed within a period of 30 years. The year of maximum growth (t_0) was 1835. After reaching its saturation level, the canal network declined rapidly due to vicious competition from railways.

---

[17] A distinction can be made with respect to the concept of "substitution" as used in economic theory. There substitution describes a case when a particular product is produced through a different combination of factor inputs, without necessarily entailing changes in technologies, processes, or techniques. Consider, for instance, an industrial boiler that can burn oil or natural gas. If prices change, oil may be substituted for gas or vice versa without requiring a new boiler or changes in industrial processes. In most cases, substitution between various factor inputs also entails changes in technologies and techniques. Thus, substitution in an economic sense, i.e., from scarce to more abundant raw materials as inputs to production, is generally impossible without technological change.

[18] For statistical measures of fit and parameter uncertainty of this and subsequent examples, see Grübler (1990a).

Box 2.1: Innovation Diffusion and Technological Substitution

The patterns and pace of the spread of innovations – in the form of new ideas and artifacts (diffusion) and the way these interact with existing ones (substitution) – are, as a rule, nonlinear. No innovation spreads instantaneously, if it spreads at all. Instead, the temporal pattern of diffusion is usually S-shaped: slow growth at the beginning, followed by accelerating and then decelerating growth, ultimately leading to saturation. The adage "Only the sky is the limit" certainly does not hold true for technologies.

As a simple and representative S-shaped diffusion/substitution curve, the logistic curve has been widely used. (Note though that the model is entirely *descriptive*, it shows how a diffusion/substitution process looks, but does not explain *why* it behaves as it does. Various causality mechanisms from learning theory to capital vintage, or turnover, models have been suggested explaining the empirically observed S-shaped diffusion/substitution patterns. In the diffusion literature, parameters of the logistic curve – like its growth rate – are linked to other explicatory economic or sociological variables such as profitability, compatibility with social norms, or even systemic variables, like complexity and size of the system being analyzed.)

The logistic curve is given by the following equation:

$$y = \frac{\mathrm{K}}{1 + \mathrm{e}^{-b(\mathrm{t}-\mathrm{t}_0)}}$$

where K denotes the upper limit (asymptote), t_0 denotes the inflection point at K/2, where growth rates reach their maximum, and b denotes the diffusion rate (the steepness of the S-curve). The diffusion rate is frequently also denoted by Δt, the time a process takes to grow from 10% to 90% of its ultimate potential K. It is related to the growth rate b by:

$$\Delta\mathrm{t} = \frac{1}{b}\log 81 = \frac{1}{b}4.39444915\ldots$$

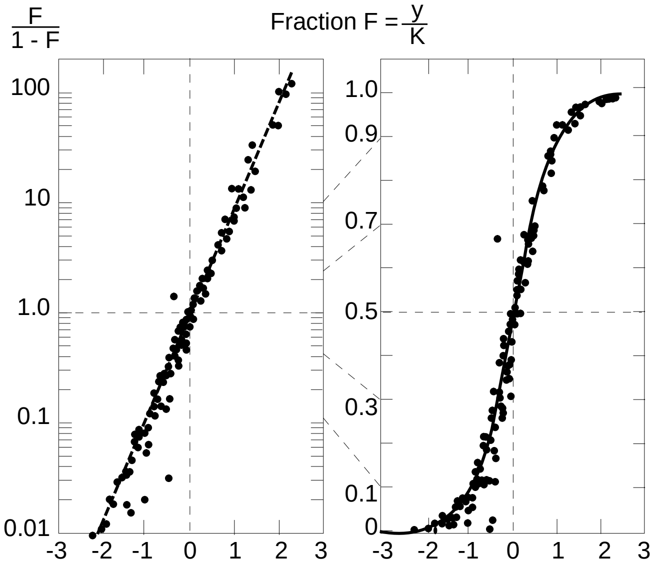
Δt also denotes the time to grow from 1% to 50% of K. Hence the entire diffusion life cycle spans $2 \times \Delta\mathrm{t}$.

The logistic curve can be rewritten with a linear right hand side, frequently used when plotting relative market shares $\mathrm{F} = y/\mathrm{K}$:

$$\log \frac{y}{\mathrm{K} - y} = b(\mathrm{t} - \mathrm{t}_0)$$

Here the interaction between the growth y achieved (or market share F), versus the growth $\mathrm{K} - y$ (market shares $1 - \mathrm{F}$) remaining to be achieved, yields a straight line when plotted on a logarithmic scale. This linearization, subsequently referred to as *logit transform*, highlights in particular the often turbulent early and late phases of the diffusion process. Note though that in this linearization zero or exactly 100% market share ($\mathrm{K} = 1$) cannot be shown.

The following graph (from the classic 1971 paper by Fisher and Pry) illustrates the life cycle in the diffusion of 17 technological innovations, measuring their relative market shares F. For simplification, the symmetrical declining shares of the older technologies being substituted are not shown. Examples of technological substitution studied by Fisher and Pry include the replacement of natural by synthetic fibers, and the replacement of traditional steel making processes by the basic oxygen process.



Statistical uncertainties of parameter estimation of logistic curves are discussed by Debecker and Modis (1994). Corresponding uncertainties and measures of goodness of fit of numerous examples are given in Grübler (1990a). As a rule, however, the human eye is an excellent guide for judging whether a particular technological diffusion or substitution path follows an S-shaped, e.g., logistic, pattern. Hence, for the sake of brevity, no curve-fitting statistics will be reported here.

Diffusion or substitution processes can also show deviant behavior from simple logistic patterns. In almost all cases this is due to the fact that a new technology, initially replacing an old technology along a logistic substitution pattern, becomes challenged by yet a newer technology, and is substituted in turn.

In the *logit transform* this shows as follows: a technology initially follows a linear diffusion/substitution pattern, that with a curvature passes through a peak significantly below the maximum possible (K = 1, i.e., 100%), in order to decline again along a linear (i.e., logistic) path. This is due to the fact that it is being substituted by yet a newer technological solution. Therefore it is quite misleading to analyze particular technologies in isolation, e.g., in the form of binary (one-to-one) substitution models. Only a holistic analysis can allow conclusions to be made on the particular shape of the diffusion/substitution trajectory technologies follow.

A generalized model for multiple competing technologies was first proposed by Marchetti and Nakićenović (1979), and some illustrative examples are given in the subsequent chapters (cf. e.g. *Figure 2.12*).

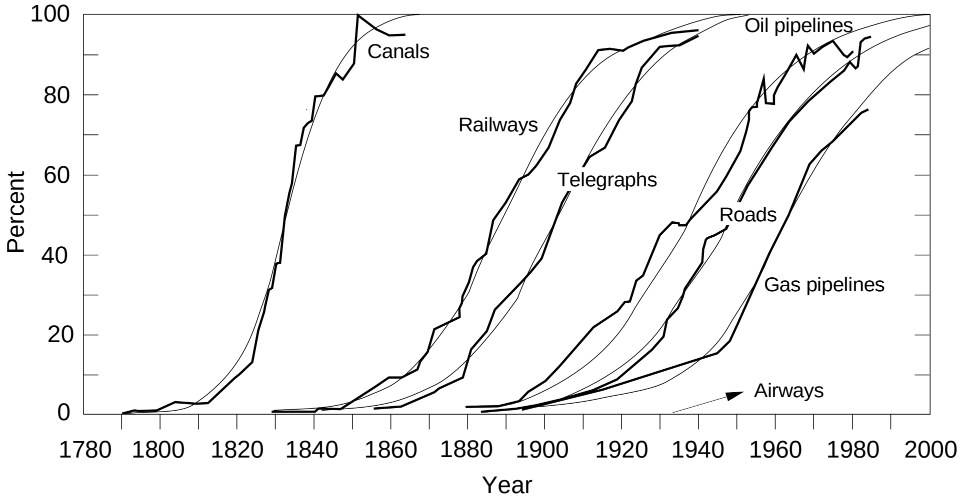


Figure 2.10: Growth of US transport infrastructures as a percentage of their maximum network size, empirical data (bold jagged lines) and model approximation (thin smooth lines). Source: Grübler and Nakićenović (1991). For the data of this graphic see the Appendix.

Figure 2.10 illustrates that subsequent transportation infrastructures, e.g., railways and roads, followed a similar pattern. In the figure the different sizes of individual networks have been renormalized to emphasize their similar diffusion patterns. The absolute saturation size of the railway network is an order of magnitude greater than that of canals. For the road networks, the saturation size is two orders of magnitude greater. Not surprisingly, their diffusion rates are slower. Δt equals 55 years for railways and 64 years for roads, compared with 31 years for canals. It is also interesting to note the regular spacing in *Figure 2.10* – about half a century between the three major historic transport infrastructures – and to note the close relationship between different infrastructures. Railways and the telegraph evolved together, as did road networks and oil pipelines necessary to transport the oil fueling the road vehicles. These examples illustrate the importance of technological interdependence and cross-enhancement, and the necessity of analyzing the diffusion of technologies in the larger context of technology "families" and "clusters".

Figure 2.11 illustrates a particularly striking case of technological substitution: the replacement of horses and carriages by cars. The figure shows the numbers of (urban) riding horses and cars in the USA and the practical disappearance of the horse as a transport technology within less than

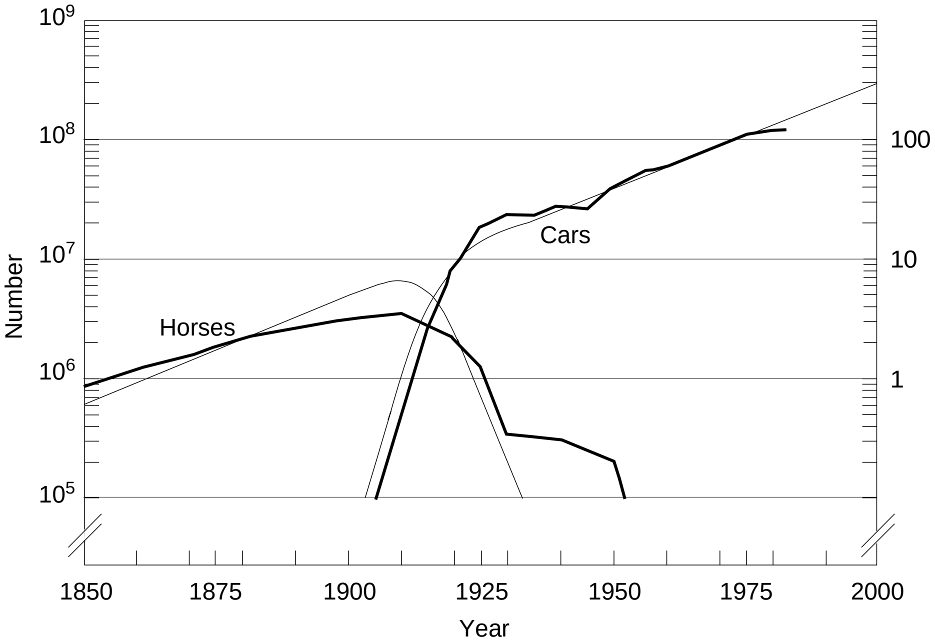


Figure 2.11: Number of (urban) draft animals (horses) and automobiles in the USA, empirical data (bold jagged lines) and estimates (thin smooth lines) from a logistic model of technological substitution. Source: Nakićenović (1986:321).

three decades. Δt equaled approximately 12 years. [The Nakićenović (1986) estimate refers to nonfarm horses only, peaking at over three million in 1910. Farm horses (many of them also used for transport purposes) totaled over 20 million in that year.] The substitution was undoubtedly fast enough to traumatize oat growers and blacksmiths, but it also created new job opportunities in gasoline stations, in the oil industry, in auto repair shops, and elsewhere.

The substitution of an old technology by a new technology shown in *Figure 2.11* is a simple example of the general case of technological change in which there are several competing technologies. *Figure 2.12* shows the introduction of the first generation of emission controls in the US automobile fleet followed later by the technology of catalytic converters. Note that the diffusion rates (Δt) in *Figure 2.12* are about 12 years, the same as that in *Figure 2.11*. This suggests that the replacement dynamics of road vehicle technologies have not changed very much. The most likely explanation is that the lifetime of road vehicles has remained relatively constant: the working lives of horses and cars are both about 10 to 12 years.

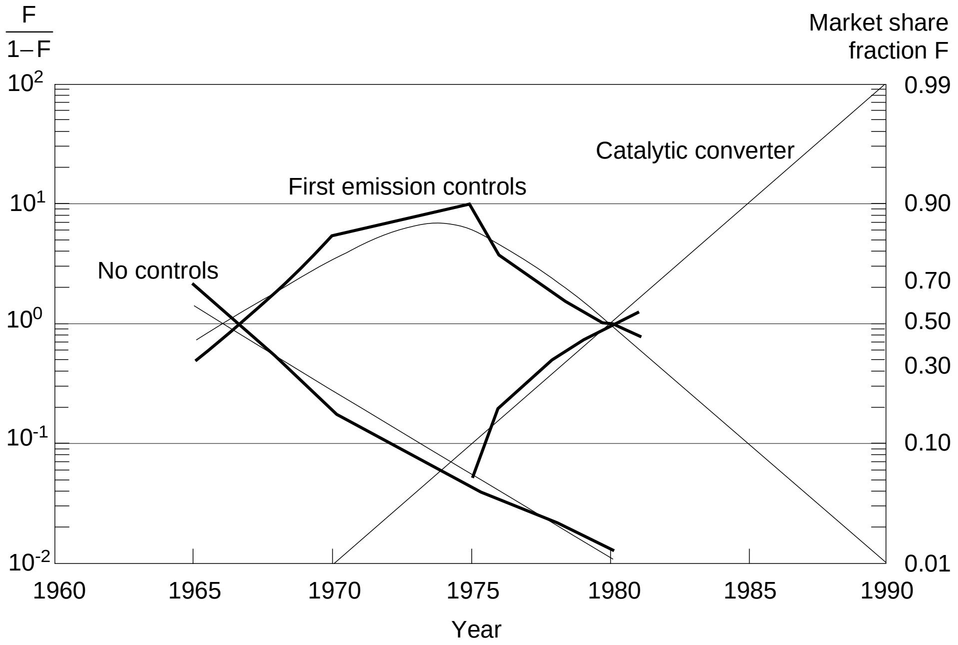


Figure 2.12: Diffusion of cars with first emission controls and catalytic converters and displacement of cars without emission controls in the USA, in fractional shares (F) of total car fleet, empirical data (bold jagged lines), and estimates (thin smooth lines) from a logistic substitution model. Source: Nakićenović (1986:332). For the data of this graphic see the Appendix.

The example of the automobile illustrates yet another dynamic feature of technological evolution: growth beyond the initial field of application. The car industry grew initially by replacing horses. That stage of its growth was completed in the 1930s. Subsequently new markets developed: long-distance travel in competition with the railways and short-distance commuting that enabled, and responded to, changing settlement patterns characterized by suburbanization. The result is approximately 135 million automobiles registered in the USA, roughly 0.6 cars per capita. As mentioned above, however, other countries will not necessarily follow an identical path. The high density of cars in the USA results from specific initial conditions including high individual mobility, even before the automobile, and from a long sustained period of diffusion that created precisely the lifestyles, spatial division of labor, and settlement patterns of an "automobile society". In short, it is yet another example of "path dependency".

Some "Stylized" Facts on Diffusion

The above brief description reiterates the main result derived from thousands of diffusion studies: no innovation spreads instantaneously. Rather, diffusion follows a very consistent pattern of slow growth at the beginning, acceleration of growth via positive feedback mechanisms, and finally saturation. Of course timing and regularity of such processes vary. But the important lesson to retain is that diffusion in most cases of any economic or social significance takes several decades. (For a comparative cross-national study of technology diffusion in industry, see Nasbeth and Ray, 1974; and Ray, 1989.) For large-scale and long-lived infrastructures it may take up to 100 years (Grübler, 1990a).

Diffusion is also a spatial phenomenon. It spreads from focused innovation centers, through a hierarchy of subcenters, to the "periphery" of diffusion (cf. Hägerstrand, 1967). *Figure 2.13* illustrates the spatial diffusion of railway networks in Europe. The construction of railway networks in England spanned approximately 100 years, while it took only half as long in Scandinavia. Railway networks were also more extensive in the countries leading the introduction of this technology (i.e., England and the USA) than in countries that followed later (*Figure 2.14*).

By 1930 the core countries in railway development (England, the rest of Europe, and the USA) had constructed 60% of the world's 1.3 million km of railways. The global railway network has not increased since then because of the introduction of newer transportation systems. These systems follow patterns that are similar to those of the railways. Automobile diffusion at the global level corroborates the accelerated diffusion rate (learning of late adopters) and their lower adoption densities (Grübler, 1990a). Thus, uneven adoption levels are likely to persist, particularly as new transport systems are developed in response to concerns over environmental impacts and changing societal needs. In the case of the automobile, we might expect alternatives to the internal combustion engine to become available within the next few decades, a development that would lead to considerably lower future energy demands than currently assumed (Grübler *et al.*, 1993b; see also Chapter 7 below).

Figure 2.15 summarizes the following main "stylized" facts representative of both theoretical and empirical diffusion research:

- No innovation spreads instantaneously. Diffusion typically follows an S-shaped temporal pattern. The basic pattern is invariant, although the regularity and timing of diffusion processes vary greatly.

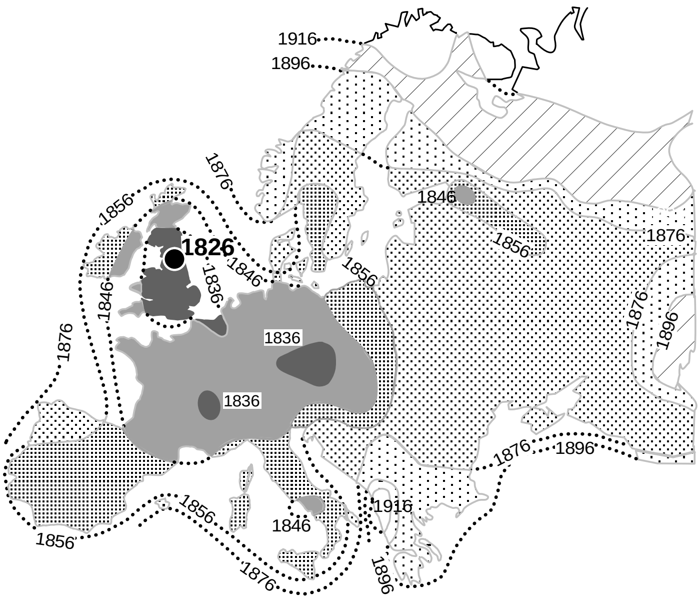


Figure 2.13: Spatial diffusion of railways in Europe, in 10 year isolines of areas covered by railway networks. Source: adapted from Godlund (1952:34).

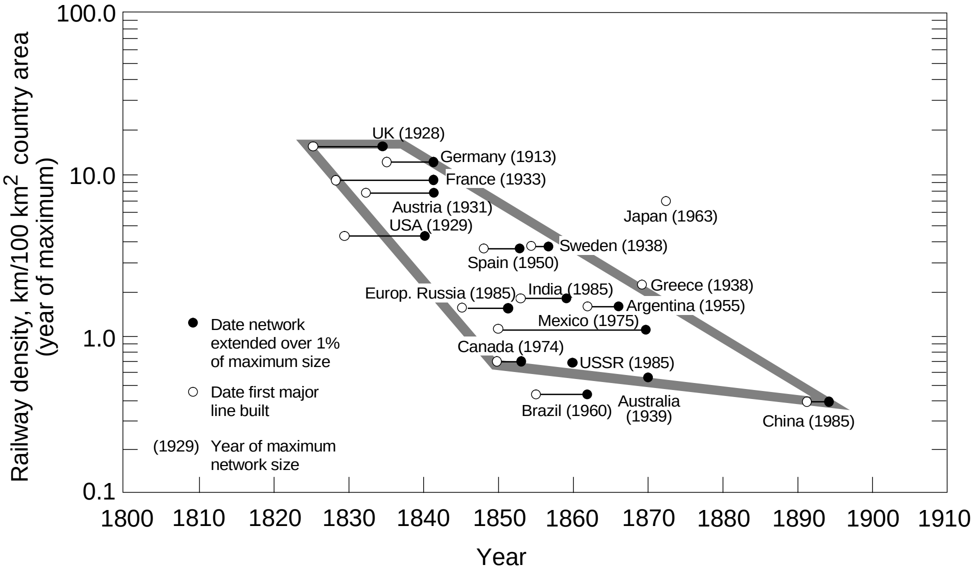


Figure 2.14: Spatial railway densities (in km railway lines per 100 km^2 country area) as a function of the introduction date of railways. Source: Grübler (1990a:98).

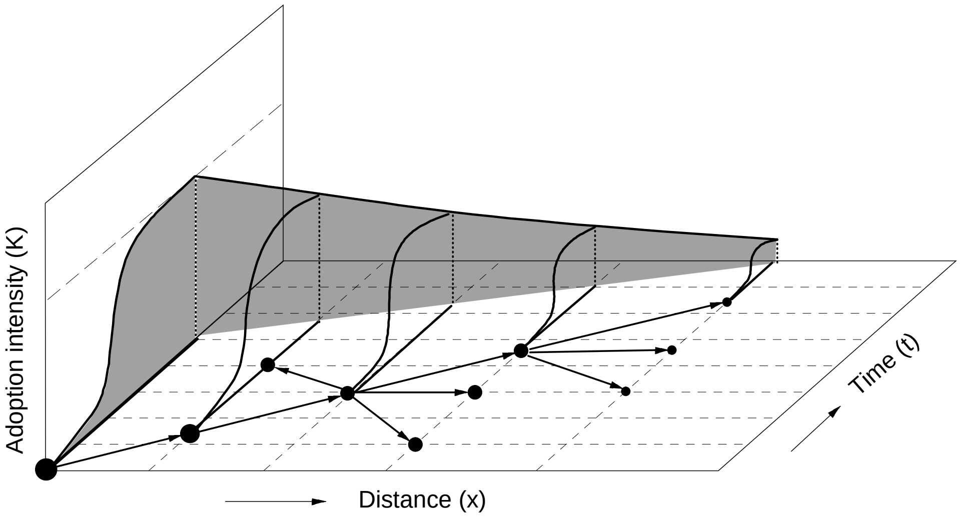


Figure 2.15: A conceptual representation of the diffusion process in time and space. Source: adapted from Morill (1968).

- Diffusion is both a temporal and spatial phenomenon. Originating in innovation centers, a particular idea, practice, or artifact spreads within a core area and then, via a hierarchy of subcenters, to the periphery.
- Although starting later, the periphery profits from the experience gained by the core and generally has faster adoption rates. Quicker adoption, however, results in a final lower adoption intensity than in core areas.
- Because of such differences, application densities and the timing of diffusion are not uniform in space, among the population of potential adopters, or across different social strata. In particular, there is little theoretical or empirical evidence to assume that adoption intensities of early diffusion starters are any guide to the adoption levels of late followers.

What governs the pace of technological diffusion? At the microlevel of the individual consumer or firm a number of factors have been identified (see e.g., Rogers, 1983):

- *The perceived relative advantage of a new artifact or technique.* This has been the focal point of diffusion studies in economics. Key variables include profitability and the required size of investments. Other things being equal, the higher the perceived profitability and the lower the required investments, the faster diffusion proceeds.

- *Compatibility.* Sociological and anthropological studies have identified compatibility with social values and with existing practices and techniques as important determinants of diffusion rates. In economics "network externalities", i.e., requirements for additional infrastructures or the existence of standards facilitating interchanges, have also been identified as important variables. For example, the diffusion of electric appliances in areas without an electricity grid is unlikely. In the early days of video recorders the existence of three different major cassette standards reduced the possibilities of sharing or renting cassettes, thereby slowing diffusion.
- *Complexity.* By complexity we refer to the learning and knowledge requirements for producing and using new artifacts and techniques. Anthropological, technological, and economic diffusion studies invariably identify complexity as an important variable. However, because quantitative measures for complexity are difficult to develop, its influence is usually described in qualitative terms.
- *Testability, observability, and appropriability.* Diffusion proceeds faster if a new artifact or technique can be tried out, if experience and information from peers is available, and if an innovation is easy to obtain. Starting with the French sociologist Gabriel Tarde in 1890, a number of research streams (e.g., Bandura, 1977) have analyzed diffusion processes primarily as learning and social imitation phenomena. In the words of a Chinese proverb, "If you want to become a good farmer, look at your neighbor". While mass media like television or the press are effective in spreading general information about an innovation, actual adoption decisions appear to be made based on interpersonal communication with peers and neighbors. It may be reassuring that today's PC users are not very different from Chinese farmers of 1,000 years ago. The fundamental lesson is that interactions within small social networks are important, take considerable time, and should not be "shortcut" through top-down centralized marketing efforts. Economic studies also emphasize the importance of informal information networks and close cooperation between buyers and suppliers, i.e., good appropriability conditions.

The macrolevel factors governing the rate of technology diffusion include, first, the size of the system involved (bigger systems entail longer diffusion time) and, second, whether the process is one of technological substitution or pure diffusion. Substitution involves replacing existing techniques or artifacts, while pure diffusion entails creating an entirely new social, economic, and spatial context, which obviously takes a longer time to achieve, or can even block diffusion in the first place. These are the macrolevel equivalents of the complexity and compatibility variables discussed above.

Although the driving forces and factors determining the speed and extent of diffusion are varied and change over time, at the macrolevel the transition paths have a very ordered structure. Diversity and complexity at the microlevel result in overall orderly transition paths, and according to recent theoretical findings (see e.g., the discussion and simulation models of Dosi *et al.*, 1986; Silverberg *et al.*, 1988; and Silverberg, 1991), such diversity appears to be even a prerequisite for diffusion.

Finally, it is important to recognize the pervasiveness of uncertainty and imperfect information in all decisions concerning technology diffusion. These factors affect the assessment of existing artifacts and practices, and more particularly, of new alternatives. Any adoption decision involves personal "technological forecasts" and varying degrees of risk aversion. Individuals, firms, and organizations cannot be modeled as economic "robots" with perfect foresight and economic "rationality". This is particularly true for the early diffusion phase of a technology, where decisions are especially complex and uncertain.

2.2.4. Technology selection: Abundance of nonstarters, uncertainty, and opposition

Any realistic history of social and technological innovation would consist mostly of "nonstarters", i.e., examples of innovations that failed to diffuse altogether. The existence of a possible solution (innovation) is therefore by no means a guarantee for subsequent diffusion. *Figure 2.16* shows an amusing failure suggested in 1828 by Henry R. Palmer – a monorail railway using sails. By then Stephenson had built his first railway line, and the dependence of Palmer's innovation on the vagaries of the winds would seem to have made for long odds. Nonetheless, it is fair to assume that the race was still far from settled at that time, and the ultimate success of the steam railways would have been very difficult to predict.

A good example of both the uncertainty in the early phases of technology development and the abundance of nonstarters is the problem of preventing dangerous smoke sparks from steam railways. Smoke sparks from wood-burning steam locomotives in the USA represented a serious fire hazard. Over 1,000 patents for "smoke-spark arresters" were registered in the 19th century (some illustrated in *Figure 2.17*) in a futile search for a solution. Ultimately none of these was successful, and the problem was solved not by an incremental "add-on" technology, but by the replacement of steam by diesel and electric power.

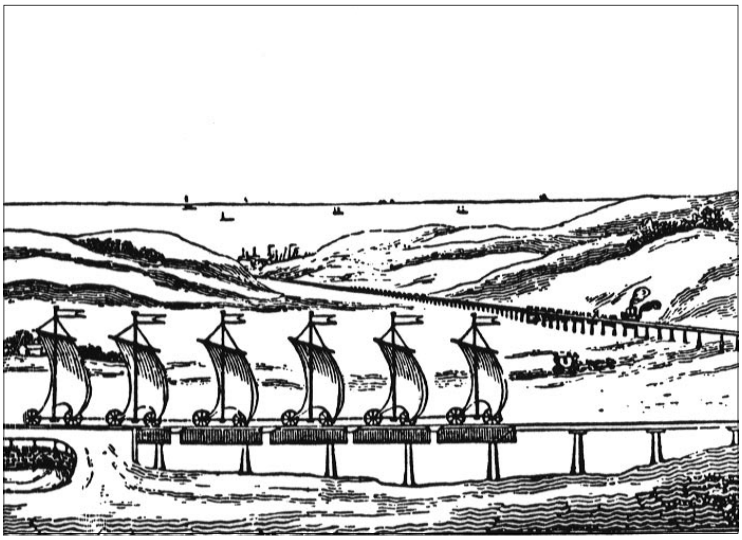


Figure 2.16: A failed innovation: monorail using sails, as proposed by Henry R. Palmer in 1828. Source: Marshall (1938:171).[4]

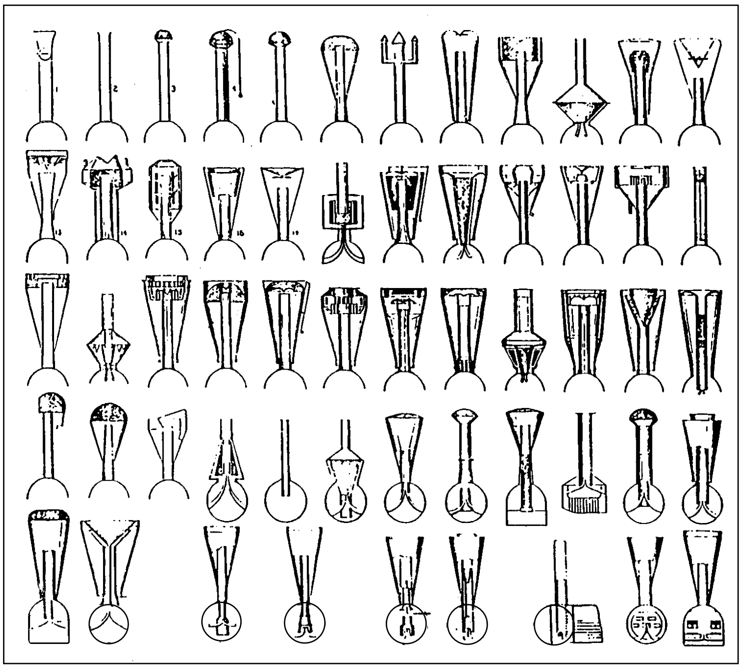


Figure 2.17: Technological variety in response to an environmental hazard. A few examples of the more than 1,000 patented “smoke-spark arresters” for wood-burning steam locomotives in the USA. Source: Basalla (1988:136).[5]

This large variety of possible alternatives illustrates the extent of the diversity and experimentation that precede successful diffusion. "Many are called, but few are chosen". Often the period of experimentation is lengthy. The current standard railway gauge in Europe emerged only after considerable time. (Spain and Russia continue to use different gauges, creating the inconvenience of changing trains at the border or changing the train bogeys.) Even in the USA, a single country, standardization of different railway gauges took several decades, as each company was reluctant to make the costly investments to retrofit their railway lines. In the case of road traffic the decision to drive on the left or the right side was also not straightforward. There were even instances where both standards prevailed at the same time.[19]

Standards are essential for technological systems to function smoothly. We can define standards simply as a set of technical specifications that assure intra- and interoperability of technologies (see *Box 2.2*). Intraoperability refers to technologies functioning within their specific infrastructures (e.g., a locomotive that can run on standard gauge railway lines). Interoperability refers to standards enabling the "exchange" between otherwise distinct technologies (e.g., standard dimensions for containers that can be loaded from a ship onto a truck, or the now ubiquitous data file transfer protocols for exchanging data between computers with different operating systems and file structures).

Optimality is of secondary importance, as any standard is better than none at all. Indeed, the issue of "bad" technology choices has received considerable attention recently, stimulated by the work of Brian Arthur (1983, 1988) and Paul David (1985). The two most prominent examples cited are the choice of the internal combustion car at the turn of the century over steam and electric alternatives (Arthur, 1988) and the choice of the QWERTY keyboard standard for typewriters (David, 1985). Arthur and David argue that both choices were inferior to the alternatives available at the time, and are therefore examples of suboptimal choice "by historical accident". They have been challenged both by economists defending the neoclassical dogma (e.g., Liebowitz and Margolis, 1990, 1995) and by historians (e.g., Kirsch, 1996). Although the steam car won a number of early automobile races, the internal combustion engine offered a much higher power to weight ratio (especially important considering the bad roads at the time) and no requirement for frequent water refilling. It also had a much larger range

---

[19] Between 1918 and 1938, the western part of Austria drove on the right side of the road, and the eastern part drove on the left. Italy in the 1920s was even more complicated; in major cities where tramways drove on the left (reflecting their origin in England), cars also drove on the left. In the countryside, cars drove on the right.

Box 2.2: Technology Standards

Technology standards are a set of codified technology characteristics that enable:

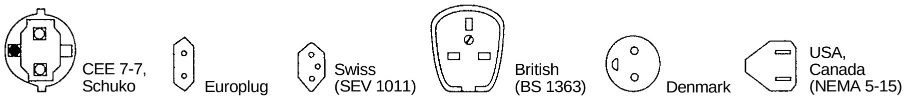
- *Interchangeability* (e.g., electricity plugs of different devices all fit into the wall sockets.
- *Product information* (e.g., producers and consumers alike can rely on standardized product qualities).
- *Interoperability* (e.g., a train can operate throughout the entire railway network if gauges are standardized).
- *Regulation* (e.g., through establishing environmental standards).

Standards can emerge spontaneously (*de facto* standardization), or can be the intentional outcome of a formal process of cooperation between companies (e.g., between different equipment manufacturers of compact discs) or of administrative procedures (*de jure* standardization).

The first wave of standardization originated at the end of the 18th century and aimed toward industrial rationalization. A typical example of this would be standardized metal construction parts that could be used for building a whole range of structures, from bridges to the Eiffel tower. The main economic rationale of such standards is the exploitation of economies of scale.

A second (and in some ways parallel) wave of standardization originated from the increasing complexity of products and the increasing size of markets. This created information asymmetry problems between sellers and buyers of products. Quality standards help to evaluate product quality without requiring costly inspection and test procedures. (For instance, at a gas station, the consumer needs to be sure that "unleaded" is indeed unleaded gasoline.) The main economic function of this type of standard is the reduction of transaction costs.

The third category of standards enables exploitation of so-called *network externalities*, where the (economic and user) value of a network (from railways to information technologies such as the telephone system) increases with its size. This requires interoperability and interconnectivity (interface or compatibility standards) among initially independent and incompatible networks that can co-exist sometimes for extended periods of time. For instance, it took nearly 50 years before the different gauges of private railway companies became standardized in the USA enabling a train to run from the east to the west coast. Spain and the former USSR continue to use a different (wider) railway gauge from the rest of Europe (and as illustrated below, a diversity of electric plugs standards still persists).



The last standardization movement is more recent: the use of standards as regulatory instruments to increase social welfare such as health, or environmental quality. Minimum quality standards or levels fix the maximum allowable levels (e.g., of emissions, noise, or of pollution and toxics in water and food). Obviously these standards change over time, influenced by increasing knowledge of negative effects and the availability of new technologies to monitor and measure ever more dilute concentrations.

Dominique Foray
Centre National de la Recherche Scientifique, Paris, France

than the electric car, an advantage that continues today. The QWERTY keyboard design is argued to be ergonometrically inferior to alternative layouts (e.g., the Dvorak design). In the age of mechanical typewriters, however, the resulting reduction in typing speed and less frequent hammer blocking may have been desirable features of the QWERTY layout.

The QWERTY keyboard is a good example of the extent to which we are often "locked-in" to particular configurations, artifacts, technological systems and standards (Arthur, 1988).[20] A particular solution that may have been best at an earlier time, but now faces superior alternatives, can often only be dislodged with great difficulty and at high costs. Not only do technologies change; so do social, environmental, and technological priorities and requirements. Given such changes, the existence of a large stock of technologies and infrastructures strongly influenced by past decisions creates formidable challenges, and can even become an obstacle for the introduction of newer systems and of economic growth (cf. the classic paper by Marvin Frankel, 1955). However, this is no real news. Societal concerns have been, and continue to be, important forces shaping technology systems. In turn, dominant technological systems are difficult to change within a short period of time.

Such challenges are not insurmountable, and indeed technologies eventually become adapted to changing social preferences. The bicycle is an example of such an adaptation and of the extent to which social fashion drives initial technological designs. Today's bicycle, with front and rear wheels of equal size, is derived from the safety bicycle design that emerged at the end of the 19th century. Its design is radically different from earlier bicycles, particularly the famous Penny-farthing (*Figure 2.18*).

Why were the Penny-farthing and (the name tells all) *Boneshaker* designs successful in the 19th century, whereas the safety bicycle only emerged at the end of the century? The answer lies in the changing expectations that people projected onto the technology. The Penny-farthing's main appeal was to "young men of means and nerve" (Pinch and Bijker, 1987:34). Such an athletic image conveyed by customers and producers alike neglected women

---

[20] Technological "lock-in" is often referred to as "path dependency". We prefer to use the term "lock-in" to describe a particular historical choice that becomes almost irreversible, standards being the most apparent example. We will use "path dependency" for describing apparent stabilities in macropatterns of technological change resulting from the accumulation of many decisions moving in a persistent direction. These are not the result of a discrete historical event or "accident". They result from persistent "signals" driving technological change in one particular direction and thereby creating irreversibilities, or at least substantial inertia. We return later to the issue of path dependency when we address theories of induced technical change.

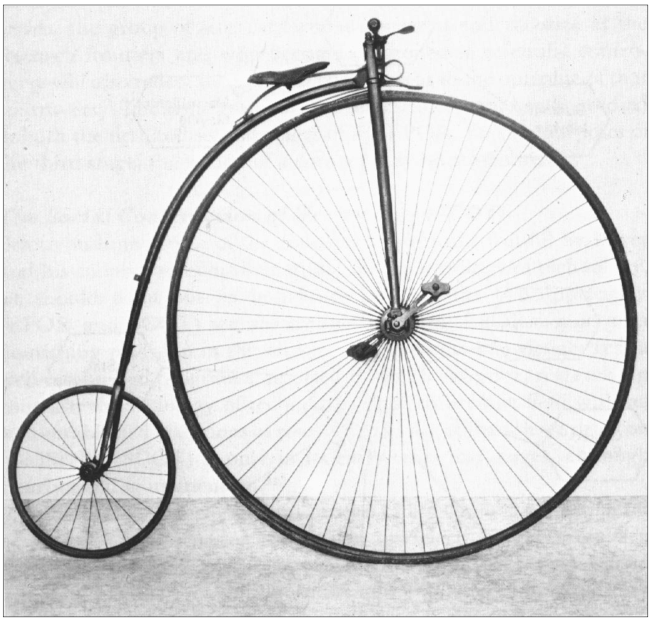


Figure 2.18: A typical Penny-farthing bicycle (Bayliss-Thomson Ordinary of 1878), a design for "young men of means and nerve". The safety bicycle design (resembling the bicycle of today) evolved much later. Its rather bumpy development history was apparently strongly influenced by the social construction of "what a bicycle had to be". Quotations from Pinch and Bijker (1987:28–34). Photograph courtesy of the Science Museum London/Science & Society Picture Library.

with their cumbersome 19th-century dress code. It took many unsuccessful design innovations, several confluent technology developments (Dunlop pneumatic tires and the rear chain drive), and 20 years before the alternative design and social image of the bicycle that we know today stabilized: a bicycle as a safe and comfortable transport device, that anybody could ride. This "social constructivist" perspective emphasizes feedbacks between consumers and designers, between actual and potential users, and among different social groups promoting or resisting particular technological configurations and designs.

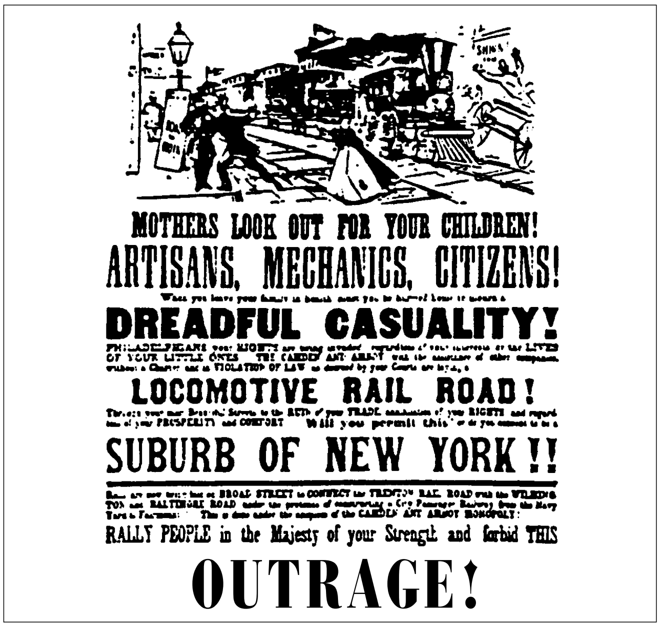
Such interactions usually pass unnoticed. They become most apparent in instances of violent opposition to technological change. Such opposition is a recurrent historical phenomenon – from the Luddites, to resistance against railway construction (*Figure 2.19*), and modern-day concerns over job losses and NIMBY (Not In My Back Yard) resistance. The Luddites were organized bands of English handicraftsmen who sought to destroy the textile machinery that was displacing them. They were named after their imaginary leader, King Ludd. The movement started in 1811 in Nottingham and spread quickly. It was halted by severe repression, culminating in a mass trial at York in 1813, with many hangings and deportations. The pattern was to be repeated later in 1830 in the resistance of the Captain Swing movement to new agricultural machinery (see *Figure 2.20*). [The best overview of resistance to technology continues to be Stern (1937:39–66).] Interestingly, the opposition to mechanical threshing machines in rural England in the 1830s also follows the classic diffusion pattern (*Figure 2.20*). The diffusion rate of about two weeks shows the effectiveness of social networks even in the absence of modern transport and communication technologies.

Opposition to technological change is a source of uncertainty, but it can also serve as an effective selection mechanism that either eliminates socially unsustainable solutions or prompts technological designs to be responsive to societal concerns. As such, opposition illustrates best the complexity of the forces driving technological change. The interplay among social groups shapes the context in which technologies evolve and can trigger an exploration for new alternatives when existing technological combinations no longer appear sustainable.

2.3. Sources of Technological Change

There are three principal sources of technological change: (i) new knowledge; (ii) improved application of knowledge, i.e., learning; and (iii) entrepreneurship and organization. All three represent "disembodied" aspects of technology regulated through social "techniques", including institutions such as universities and R&D laboratories, media such as scientific and applied journals, and incentive systems such as patent protection. New developments in these disembodied (software) aspects of technology need to occur before embodied (hardware) technological change can take place, although embodied technological change can then lead to further advances in knowledge. New scientific knowledge leads to new technologies, but science also depends on technologies for measurements, experiments, and disseminating new knowledge. Thus, there is no simple one-way street between science and technology, or between technology (instruments, new observation technologies) and science, as convincingly argued by Adams (1995:32–33).

Galileo's discovery of Jupiter's moons and his challenge of the Aristotelian dogma of the sun revolving around Earth were made possible by a new technological artifact from the Netherlands: the telescope. In turn, new



MOTHERS LOOK OUT FOR YOUR CHILDREN!

ARTISANS, MECHANICS, CITIZENS!

DREADFUL CASUALITY!

LOCOMOTIVE RAIL ROAD!

SUBURB OF NEW YORK!!

RALLY PEOPLE in the Majesty of your Strengt and forbid THIS

OUTRAGE!

Figure 2.19: Resistance to US railways: January 1838. Source: Grübler (1990a:105), courtesy of Metro-North Commuter Railroad, New York.

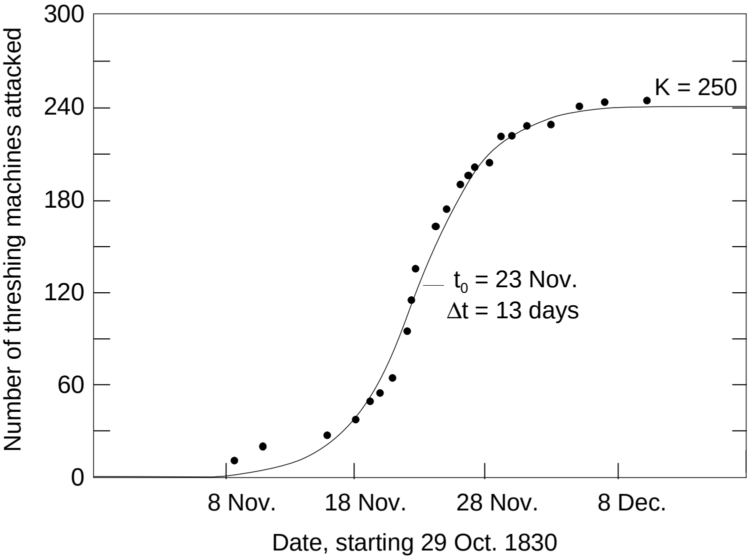


Figure 2.20: Resistance to technology as a diffusion process: number of threshing machines attacked during the Captain Swing movement in 1830. Data source: Hobsbawn and Rudé (1968:Appendix III:1–24).

scientific theories (astronomical in this case) directed and guided the further development of observational and measurement technologies. The spectrum of signals analyzed by astronomers today far extends beyond the human eye.

Knowledge takes many forms and comes from many sources. Over the past 300 years, science has emerged as the principal "technology" for generating new knowledge. Distinctions are commonly made between basic and applied science and research, and between public knowledge, proprietary knowledge, and truly private or tacit knowledge. Public knowledge is what anyone can acquire, e.g., by reading *Nature* or *Science*, or other information in the open literature. Proprietary knowledge is protected by patents and access is limited through licensing arrangements. Private or tacit knowledge includes special "tricks" in manufacturing that are largely unrecorded, known only to experienced workers and passed on largely through "hands-on" experience. There is a correlation between the institutional source of knowledge and its *appropriability.* Scientific knowledge is largely a public good, and much applied knowledge is either proprietary or tacit.

The primary institutions of science – universities, learned societies, and academies of science – date from the Age of Enlightenment, but professional and industrial R&D is a relatively recent phenomenon (Rosenberg, 1991). The first R&D labs were created for elementary tasks such as measurement and quality control. Typical first applications were measuring the metal content in ores and measuring the quality of metallurgical products. Another early application was research on possible uses for "by-products" of petroleum refining in the production of illuminating oil (Rosenberg, 1991). These early "by-products" are now principal products of oil refining: motor fuels, petrochemical feedstocks, and lubricants. Only at a much later stage did industrial R&D labs move into process and product innovation.

The distinction between basic and applied science and the development of many technologies from scientific results suggests a linear model of technological change. This model is a more detailed stage representation of the life cycle typology *invention, innovation*, and *diffusion* discussed previously. The stages of this model are as follows:

- Basic research produces new scientific knowledge (discoveries).
- Applied research leads to proposed applications (patents).
- Further applied research and development refines this knowledge sufficiently to justify substantial investments in new technology (development).
- Investments are made in new production facilities, equipment, and specific products (innovation).

- Experience leads to improvements and adaptation in early applications (early commercialization).
- Widespread commercialization leads to new levels of technical standards, economic performance, and productivity (diffusion).

To these stages we could add three more:

- Experience, learning, and feedbacks from customers lead to further technological and economic improvements and to wider fields of application.
- Pervasive diffusion leads to macroeconomic, social, and environmental impacts.
- Such impacts lead to scientific research and new information on causes of and possible solutions to adverse impacts.

This takes us back to square one, and the whole sequence starts again. Following these steps in the order just presented represents a science or technology "push" view of technological change. Were we to follow essentially the same steps but in the reverse order, we would have a "demand pull" view of technological change. Both are extreme perspectives. The first views technology development as driven exclusively by opportunities; the second views it as driven exclusively by needs.

Both linear models have been largely dismissed in the literature (see e.g., Mowery and Rosenberg, 1979, or the review article by Freemann, 1994) in favor of models with multiple feedbacks and various factors driving different phases of a technology's life cycle. In early phases science/technology push factors may dominate, whereas in later phases demand pull factors may be more important (see e.g., the work of Walsh, 1984; and Fleck, 1988).

There are certainly examples of a linear development sequence where "science discovers and technology applies", e.g., nuclear energy and the transistor and semiconductor. But counterexamples also abound. The first steam engines were built without much understanding of thermodynamics, which was developed only much later. The Wright brothers flew propelled, heavier-than-air machines, even while some physicists still proclaimed this to be impossible (Rosegger, 1996:4). Aviation developed in the 1920s and 1930s without the knowledge and technology to fly in difficult weather conditions or at night. Radar, today considered essential for aircraft navigation, was not developed until World War II.[21] Such examples emphasize the inadequacy

---

[21] The eminent sociologist of invention S. Colum Gilfillan (1935) listed 25 different technological means to overcome the limitations that fog and similar bad weather conditions represented for aviation (NRC, 1937). None of these eventually contributed toward the solution that was provided by radar. But Gilfillan was right in predicting "quite confidently" that the problem of fog would soon be overcome, and he was justified in exploring scenarios of industry development that assumed no danger from fog.

of models in which technological change proceeds linearly with "neat" divisions between science and technology. This does not create problems for scientists doing basic research within industry (AT&T scientists have, for example, been pioneers in atmospheric chemistry and the discovery of cosmic background radiation). But it can embarrass modelers who treat knowledge generation and improvements in a technology's application as exogenous to the economic system.

2.3.1. Who performs and who pays for knowledge generation (R&D)?

Tables 2.4 and *2.5* present a statistical overview of the R&D enterprise in the USA, the country with the largest R&D expenditures. Some US$160 billion were spent on R&D in 1993, about 2.5% of the gross domestic product (GDP). This is similar to the percentage of GDP spent on R&D in most of the advanced industrial economies. About two-thirds of R&D expenditures are devoted to (expensive) development work, 25% to applied research, and about 15% to basic research. By far the largest part of this research effort (70% or US$112 billion) is performed by industry, simply because it is industry that typically does development work, and development dominates R&D expenditures. Overall, industry provides slightly more than half of the total R&D funding in the USA. The role of government and other nonindustry institutions in R&D is also very important. It is justified first by the fact that much of new knowledge produced by research, especially basic research, is a public good. Nonindustrial R&D is also justified by the potentially very long lead times between the generation of new knowledge and its possible applications and the fact that new knowledge may never produce any direct economic "spin-offs". For these reasons firms are likely to underinvest in R&D that would be beneficial to society. Public expenditure in research is justified because society must consider the long-term future more than firms and value the noneconomic social and cultural spin-offs and new knowledge simply for its own sake.

Quantitative statistics, such as R&D expenditures or R&D personnel, only measure the inputs to knowledge generation. Outputs are even harder to quantify in the aggregate. Where attempts have been made to measure the R&D output of corporations, in terms of new products, improved production methods, etc., the results indicate significant economic returns to R&D. Frosch (1996:27) for instance, reports (internal) rates of return from 38% to 70% for the R&D operations of companies such as General Electric or General Motors, respectively.

Table 2.4: R&D activities in the USA in 1993, by institutional sector.

	Basic research		Applied research		Development	
Sector	Mill. US$	%	Mill. US$	%	Mill. US$	%
Federal government	2,900	11.1	4,900	12.3	8,800	9.3
Industry	4,700	17.9	26,500	66.8	81,100	85.5
Universities and colleges	16,350	62.4	6,360	16.0	3,140	3.3
Nonprofit institutions	2,270	8.6	1,920	4.9	1,810	1.9
Total	26,220	100.0	39,680	100.0	94,850	100.0

Source: National Science Board (1993).

Table 2.5: R&D funders and performers in the USA in 1993 (in million dollars).

	Source of funds						
R&D performer	Federal gov.	Non-fed. gov.	Industry	Uni. & colleges	Nonprofit inst.	Total	%
Federal gov.	16,600					16,600	10.3
Industry	31,000		81,300			112,300	70.0
Universities and colleges	16,700	1,850	1,500	4,150	1,650	25,850	16.0
Nonprofit institutions	3,700		750		1,550	6,000	3.7
Total	68,000	1,850	83,550	4,150	3,200	160,750	
%	42.3	1.1	52.0	2.6	2.0		100.0

Abbreviations: gov., government; Uni., University; inst., institutions.
Source: National Science Board (1993).

Still, in as far as the main output of R&D is new knowledge, or rather *new combinations* of knowledge, that can subsequently be applied in production (where economic returns accrue), it is indeed a formidable challenge to try to measure R&D "output" directly. Unlike measuring the capital intensity, or the energy intensity of an economic sector or industry, it is extremely difficult to measure "knowledge intensity" (Smith, 1995). Patent statistics suffer two weaknesses. Not all new knowledge is patented, and not all patented information is used. Nevertheless patent research has identified patterns of inventive activities (e.g., Pavitt, 1984) that provide useful insights into important sectoral and industry differences in knowledge generation and innovation.

Tables 2.4 and *2.5* indicate that R&D extends well beyond government-sponsored basic research and should therefore not be treated as "external" to economic activities. On the contrary, knowledge generation and technological development are an integral part of economic activity and constitute the single most important "input" to growth in a modern economy. Such an endogenous view of knowledge generation becomes even more important when analyzing improvements in technological applications as reflected in "learning curves".

2.3.2. Learning

The performance and productivity of technologies typically increase substantially as organizations and individuals gain experience with them. Such improvements reflect organizational and individual learning. Learning can originate from many sources. It can originate from "outside" an organization – an example is a company that, in order to facilitate its own introduction of a new process technology, hires a production engineer from a competitor that has already done so. Or learning can originate from the "inside" through R&D and investments in new technologies. Learning can come through improving "know-how", i.e., learning how to "make things better" with the "things" (artifacts, designs, practices, jobs, etc.) basically unaltered. Or learning can come through improving design features and economies of scale, i.e., reducing costs by building and using larger and larger units. There is, however, one strict precondition for learning. It requires effort and the actual accumulation of experience. It does not come as a free good.

Technological learning phenomena – long studied in human psychology – were first described for the aircraft industry by Wright (1936), who reported that unit labor costs in air-frame manufacturing declined significantly with accumulated experience. Technological learning has since been analyzed for manufacturing and service activities ranging from aircraft, ships, refined petroleum products, petrochemicals, steam and gas turbines, even broiler chicken. Applications of learning models have ranged from success rates of new surgical procedures to productivity in kibbutz farming and nuclear plant operation reliability (Argote and Epple, 1990). In economics, "learning by doing" and "learning by using" have been highlighted since the early 1960s (see e.g., Arrow, 1962; and Rosenberg, 1982). Detailed studies of learning track the many different sources and mechanisms (for a succinct discussion of "who learns what?", see Cantley and Sahal, 1980). Here we focus on the productivity gains from learning, and these can be very large indeed. During the first year of production of World War II Liberty ships, for example, the average number of labor hours required to produce a ship decreased by 45%, and the average time decreased by 75%. There are also cases, however, where no learning is evident, and we briefly discuss the reasons for such learning failures.

Learning phenomena are described in the form of “learning” or “experience” curves, where typically the unit costs of production decrease at a decreasing rate. Unit costs decrease along an exponential decay function. Because learning depends on the actual accumulation of experience and not just on the passage of time, learning or experience curves are generally described in the form of a power function where unit costs depend on cumulative experience, usually measured as cumulative output:

$$y = ax^{-b},$$

where y is the unit labor requirement or cost of the xth unit, a is the labor requirement or cost associated with the first unit, and b is a parameter measuring the extent of learning, i.e., the unit labor or cost reductions for each doubling of cumulative output. The resulting exponential decay function is frequently plotted with logarithmically scaled axes so it becomes a straight line (see *Figure 2.21*). Because each successive doubling takes longer, such straight line plots should not be misunderstood to mean “linear” progress that can be maintained indefinitely. Over time, cost reductions become smaller and smaller as each doubling requires more production volume, and the potential for cost reductions becomes increasingly exhausted as the technology matures.

Figure 2.21 plots the costs per kW as a function of total cumulative installed capacity for several electricity generation technologies. The figure shows how costs drop as experience accumulates. The learning curve patterns shown in *Figure 2.21* illustrate several general features characteristic of technological learning.

First, the learning rates, at about a 20% reduction in specific investment costs for each doubling of cumulative output, are quite similar across the three technologies of wind, gas turbines, and PV cells. This is true despite the initial costs of PV cells being ten times higher than the costs of gas and wind turbines. The learning rates are also similar between countries as shown by the PV costs in the USA and Japan.

Second, when costs are plotted as a function of accumulated experience rather than time, it is easier to draw useful analogies. For example, *Figure 2.21* shows that the dynamics of cost reductions for windmills in the USA in the 1980s are quite similar to those for gas turbines in the early 1960s.

Finally, note the two distinct phases of cost reductions in the case of gas turbines. There is an early rapid phase associated with R&D and technical demonstration (in the innovation phase), followed by distinctly slower cost reductions during commercialization (the diffusion phase). This illustrates

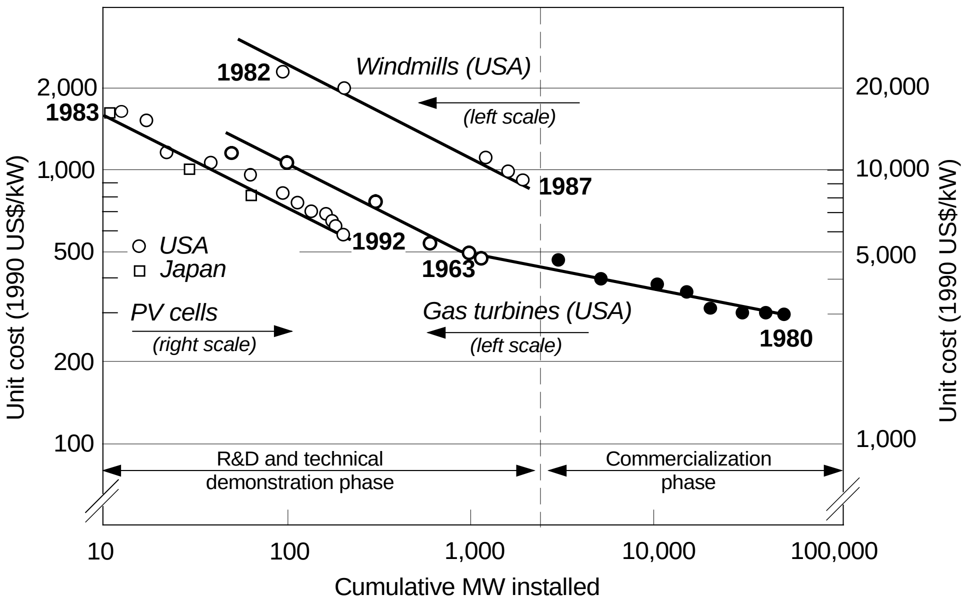


Figure 2.21: Technology learning curves: unit cost (US$ per kW) versus cumulative experience, measured by output (installed MW) for photovoltaics (right hand side scale), wind and gas turbines (left hand side scale). Note in particular the similar slope of the learning curves of the three technologies and that photovoltaics start off at costs ten times higher than the two turbine examples. Source: IIASA–WEC (1995:29).

important differences in the sources of technological learning in different phases of a technology's life cycle. As a rule, cost reductions are most substantial in early phases where R&D and design improvements yield the largest return on investments, even though benefits may not accrue directly to investors. Later entrants have the benefit of "external" learning from the improvements achieved by the "internal" learning financed by early innovators. New technological knowledge is costly to produce, but cheap to imitate. To limit external learning, or "free-riding", and to protect R&D performers, regulatory measures, particularly the patent system, have been created. Such protection is far from perfect, however. Information, learning, and experience can leak out through staff turnover, key R&D personnel being hired elsewhere, or through straightforward espionage. However, such "leakage" may be socially desirable – leading to fast diffusion of new knowledge – even if it may not be desirable for the individual firm.

The rate of learning and experience can vary enormously among different sectors and technologies. *Figure 2.22* illustrates the range of learning rates

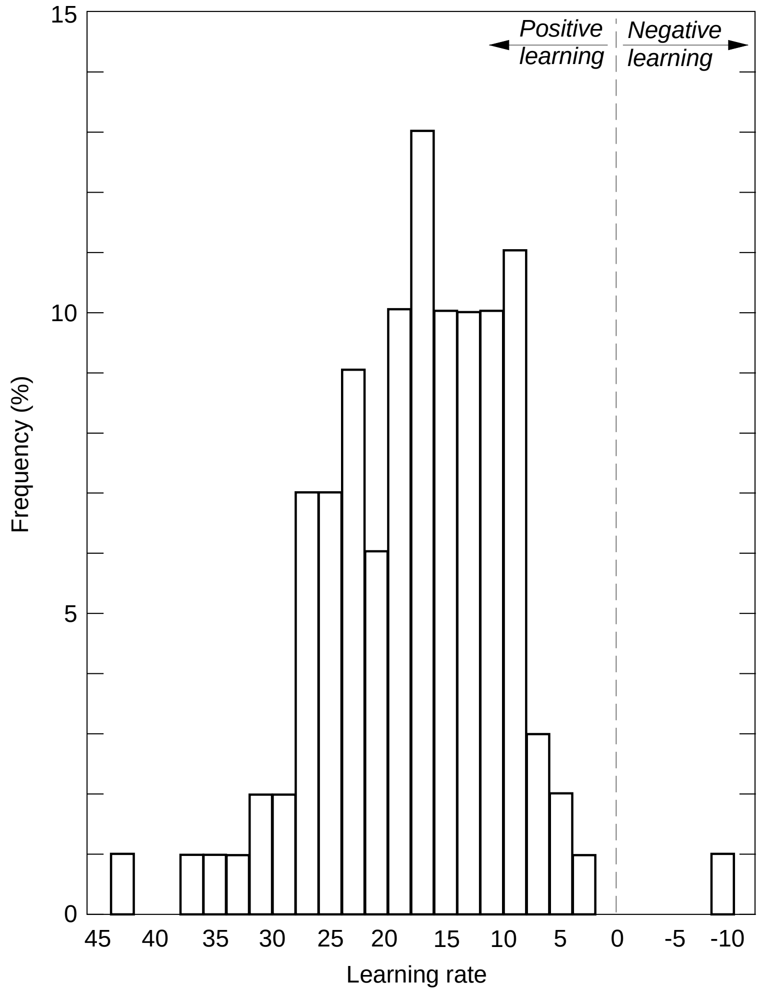


Figure 2.22: Distribution of learning rates (unit costs reduction, in %, for each doubling of cumulative output) for a sample of 108 technologies synthesized from 22 field studies. Source: adapted from Argote and Epple (1990:921).

(cost reductions per doubling of cumulative output, i.e., the parameter b in the previous equation) from a sample of 108 different technologies and products. Learning rates range from a high of 45% to only a few percent. There are also examples of negative learning, or "organizational forgetting", where costs increase rather than decrease.

In addition to learning via R&D and actual experience (investments), significant learning takes place through large-scale production. We divide large-scale production learning into three classes (see also Cantley and Sahal, 1980). These three classes are listed as follows:

1. Learning by upscaling production units (e.g., the examples of steel converters and steam turbines given previously).
2. Learning through consecutive repetition or mass production (e.g., the Model T Ford).
3. Learning through both increasing scale and consecutive repetition, referred to here as "continuous operation", i.e., the mass production of standardized commodities in plants of increasing size. The best examples are base chemicals such as ethylene or PVC (polyvinylchloride), where cost reductions have been particularly spectacular (Clair, 1983).

Such large-scale production learning usually begins at the individual plant level, but later spills over to other plants (for which this represents a source of external learning) and eventually spreads to an entire industry.

A statistical analysis of learning rates across many technologies and products (Christiansson, 1995) confirms the value of the above taxonomy and concludes that learning rates are typically twice as high for "continuous operation" as for either upscaling or mass production alone. (The mean learning rate for continuous operation in the Christiansson sample was 22%, compared to 13% for upscaling and 17% for mass production.)

A learning rate of 20% is a representative mean value advanced in the literature (Argote and Epple, 1990). Twenty percent is also the mode of the distribution function shown in *Figure 2.22.*

The example of negative learning shown in *Figure 2.22* deserves some elaboration. The example comes from the Lockheed L-1011 Tristar aircraft production. Production started in 1972 and reached 41 units in 1974. It subsequently dropped to 6 units in 1977, and increased again thereafter. The drastic reduction in output led to large-scale layoffs. When production increased again, new personnel were hired, and the experience gained initially was lost with the staff turnover. As a result, production cost reductions could not be maintained, and the planes built in the early 1980s were in real terms (after inflation) more expensive than in the early 1970s.

Thus, stop-and-go operations in R&D, and "hire and fire" strategies in production, seem to be detrimental to technological learning. Continuity in effort, in accumulation of experience, and the maintenance of human know-how seem essential for technological learning. The converse of "learning by doing" is "forgetting by not doing". This holds for R&D and production alike. Massive technology crash programs that are abandoned after a few years (e.g., the multi-billion dollar US synthetic fuel program), "stop-and-go" production schedules (e.g., of the Lockheed Tristar), or frequent design changes at considerable cost (e.g., in nuclear reactors to improve safety features) all illustrate that learning and cost reductions are not always related to scale of effort. It also depends on how efforts are organized and on the continuity and commitment of the effort. Technological "forgetting", or creating conditions not conducive to learning, can sometimes be as powerful as "learning".

2.3.3. Entrepreneurship and organization

We have discussed R&D and learning as important sources of technological change. None of these activities can take place without dedicated human effort, and it is therefore important to conclude this chapter by mentioning the human and organizational factors in technological change. These factors were particularly stressed by Joseph A. Schumpeter. He believed that the organizational entity bringing about new technological "combinations" is the firm, and that innovative activities usually do not arise out of existing firms. "It is not the owners of stagecoaches who build railways" (Schumpeter, 1911:66).[22] The creation of such firms and the promotion of particular new "combinations" was the domain of Schumpeter's "entrepreneur". Schumpeter's emphasis on the entrepreneur as the bearer of change seems to have been unduly influenced by the writings of Nietzche and prominent capitalistic entrepreneurs such as Vanderbilt, Carnegie, Edison, and Rockefeller. Schumpeter later acknowledged the importance of large organizations in performing R&D. A development engineer in the R&D department of a large electrical firm would equally qualify as a Schumpeterian "entrepreneur" (Freeman, 1994), as would a manager keen to introduce a new production process, or a marketing salesperson (a "change agent" in the terminology of the diffusion literature, see Rogers, 1983) promoting a new product.

As an example, the now ubiquitous yellow "post-it" notes were originally conceived by a 3M company employee who sang in a choir and was annoyed that the paper slips used to mark the hymns kept slipping away. The technological ingredients that were combined in post-it notes already existed; the innovation consisted of creating the new combination. The prototype, however, fell flat. Major office supply distributors thought it was silly; market surveys were negative. The product, which is now a US$100 million plus business for 3M, eventually succeeded because 3M's secretarial staff liked to use the specimens available within the company. The eventual breakthrough came with a mailing of product samples to *Fortune 500*

---

[22] A more contemporary quote in the same spirit is attributed to C.F. Kettering, the founder, and patron saint, of the GM research labs: "Never put a new technology in an old Division" (as observed by an anonymous reviewer of this manuscript).

CEO executive secretaries under the letterhead of the 3M executive secretary (Peters, 1986). While post-it notes may not classify as a major technological innovation, they are certainly a major "entrepreneurial" innovation – realized, promoted, and brought to success by individuals within a large corporation.

Such individualistic conceptions of technological change may appear naive in the age of large multinational corporations, institutionalized R&D and "big science" (de Solla-Price, 1963). But they point to the importance of organizational and institutional factors in the promotion of, or opposition to, technological change. Organizations and institutions represent social "techniques" to organize and to regulate individual human actions.

For instance, large corporations do not usually entrust the development and commercialization of new innovations to departments responsible for the existing, dominant technology. For promotion of rapid development, organizational "offsprings", such as "skunkworks", largely liberated from bureaucratic routines and tedious accountancy, have become an accepted organizational strategy. The US Army asked Lockheed in 1943 to design a new fighter aircraft, stipulating that the prototype must be delivered within 180 days. Lockheed entrusted the task to Clarence L. "Kelly" Johnson, who drew together a small team of designers, engineers and shop mechanics. They were located in temporary quarters in California near a foul-smelling industrial site, hence the name "skunkworks". [Another, or perhaps complimentary, explanation for the word comes from a popular comic strip (Lil Abner), where two brothers produce mysterious elixirs in their "skunkworks".] Johnson had 14 management rules that assured considerable informality, autonomy, and flexibility. The prototype fighter was ready in just 137 days. It was the first US jet fighter aircraft. Later technological marvels of Johnson's skunkworks were the U2 spy plane and the famous SR71 "blackbird" aircraft, which has held the speed record for air-breathing aircraft since 1962. For an autobiography of Kelly Johnson (1910–1990), see Johnson and Smith (1985).

Of course innovations continue to be created by individuals and small firms, even if the latter – if successful – do not necessarily stay "small" for long. Much has been written on the impact of firm size on innovations and their diffusion. The conclusion is that "bigger" is not necessarily "better". Internal organization within large firms is as important to innovation and diffusion as is the role of small enterprises.

New actors appear increasingly on the scene. Government-sponsored agricultural research institutions and dissemination efforts have been instrumental in introducing new crops and farming practices in the USA. Networks of institutions rather than "monolithic" R&D organizations have emerged.

The Consultative Group for International Agricultural Research (CGIAR), for example, is a network of 17 agricultural research institutions. It conducts primary research on crops and exchange of genetic resources, and also plays a major role in the diffusion of new, high-yield strains to farmers, particularly in the tropics. Environmental NGOs play an increasing role not only in opposing certain technologies, but also in actively promoting more environmentally compatible innovations. Greenpeace Germany, for example, commissioned a small company (Freon) in the former German Democratic Republic to design a refrigerator without ozone-depleting CFCs. The successful design forced all major refrigerator companies to quickly offer CFC-free models also (much to the detriment of the small, innovative company).

Thus, the portfolio of change agents is larger than ever, and their motivations, incentives, risk perceptions, and views of the future are ever more diverse. The notion of a single representative "agent" of technological change is outdated, although it continues to be used in much of the mainstream modeling of technological change, as discussed in the following chapter.

Finally, it is important to dismiss the notion of "lonely heroes" as innovators and agents of technological change. People communicate with each other, exchange ideas and information, and thereby create joint "technological expectations" (Rosenberg, 1982). These influence the visions, missions, and expectations of all those involved in research and development, marketing, etc. Because everybody expects things will develop in a particular direction, research and development focus on that direction. The model of the self-fulfilling prophecy is entirely appropriate here. It has been shown that joint expectations in the microchips business, expressed in shared technological forecasts,[23] helped establish targets, drive research, and achieve results in line with the motivating expectations (Mackenzie, 1991; Benzoni, 1992). Motivating expectations also encompass consumers. Those in the market for personal computers, for example, time their purchases based on shared expectations that prices will inevitably drop and that the next generation of models will be more powerful and their performance will be better than their forebears.

---

[23] Gordon E. Moore, Director of Fairchild Semiconductors (and one of the co-founders of Intel Corporation), postulated in 1964 that, based on trends since 1959, the number of transistors per integrated circuit would double every year or so (Benzoni, 1992:25). By mid-1995 the number of transistors per chip had reached about 100 million, basically on track with "Moore's law".

Copyright acknowledgments

[1] Reprinted from Rosegger, G. *The Economics of Production and Innovation: An Industrial Perspective*, 3rd edition, p. 175. Copyright © 1996, with kind permission from Butterworth-Heinemann.

[2] Reprinted from *Omega – The International Journal of Management Science*, Volume 15, Number 2. Rosegger, G., and Baird, R.N. "Entry and exit of makes in the automobile industry, 1895–1960: An international comparison", p. 96. Copyright © 1987, with kind permission from Elsevier Science Ltd.

[3] Reprinted from Hugill, P.J. *World Trade Since 1431: Geography, Technology, and Capitalism,* p. 256. Copyright © 1993, with kind permission from The Johns Hopkins University Press.

[4] Reprinted from Marshall, C.F.D. *A History of British Railways Down to the Year 1830,* p. 171. Copyright © 1938, with kind permission from Oxford University Press.

[5] Reprinted from Basalla, G. *The Evolution of Technology,* p. 136. Copyright © 1988, with kind permission from Cambridge University Press.