Part IV Innovation and the evolution of firms

Preface to Part IV

Richard R. Nelson
Columbia University, New York

Over the past century the locus of inventive activity has shifted from the workshops of individuals to the organized laboratories of business firms. There are several factors behind the shift. First, in many if not all industries inventing has become an activity requiring special skills and training at an advanced level in science and engineering, the use of expensive instruments and other equipment, and often a group of people working collaboratively. Second, knowledge about prevailing technology and its strengths and weaknesses, and about user-needs, has tended to become proprietary, not readily accessible to someone not connected with a firm and lacking special access. Third, invention has become a prominent component of the competitive weaponry of firms.

In recognition of the central importance of industrial R & D, if not always with understanding of the factors behind it, for many years economists have concerned themselves with the character of firms and of industry structures that are conducive to industrial innovation. The questions explored have included prominently the following. Does modern industrial invention and innovation require the presence of large and well-established firms? What are the circumstances in which small firms, or new firms, are the principal sources of industrial innovation, rather than large established firms? Is an industry structure in which only a few firms control a large share of the market, and are relatively immune from short-run liquidity crises, conducive to innovation? How to explain the instances in which rapid technical change was created through an industry structure where entry is easy and external financial sources inclined to bet heavily on promising ideas? These kinds of questions have been on the research agenda of economists at least since the publication of Joseph Schumpeter's *Capitalism, Socialism and Democracy* in 1942.

As the kinds of questions indicate, until recently the focus of economic research has been on how market structure influences industrial invention. There also has been considerable attention to the role of demand, what users will buy and how much they will pay for it, in influencing the kinds of technical advances firms bring forth. However, in recent years certain new or strengthened understandings have significantly changed the research agenda, adding new questions, and changing the way older ones are posed. Scholars have come to recognize better that in many industries the process of technological advance has a strong internal logic, which influences what demands can and cannot be met. Also, it may be much easier for firms to appropriate the returns to certain kinds of inventions than to others. These understandings have enriched, but also made more complex, analysis of the forces influencing the rate and direction of industrial invention.

In many cases it is apparent that scientists and technologists clearly see certain directions, avenues, down which they believe technology can be advanced relatively reliably, while attempts to advance the technology in other directions may be much more problematic. Where these easily pursued directions correspond to user-needs, and where innovators have some mechanism for assuring that they receive a non-trivial fraction of the use-value of their innovations, technological change tends to proceed down those tracks.

This new understanding that technology tends to unfold in particular directions in turn led to awareness that firm behavior, and industry structure, may be molded by the way technology is unfolding, at least as much as the character of innovation depends on firm behavior and market structure. The causal links go both ways, not just in one. Thus in recent years the natural ways to advance recombinant DNA technology have, in general, not required massive resources and giant laboratories, but have been pursuable by small companies, or even by individuals with access to modern laboratory equipment. And judicial decisions regarding patentability have made it possible for small-scale innovators to hold off large-scale potential imitators, and to be in a strong position regarding bargaining about patent rights. In contrast, the advance of modern passenger-jet aircraft technology has required large-scale research and development efforts on the part of teams of experienced scientists and engineers. Further, jet engine and air-frame design, while not protectable by patents, is very difficult to reverse-engineer. This is a context in which large established firms have a major advantage over newcomers.

The chapters in this section develop various of these themes. Dosi's chapter elaborates the discussion above, and presents a broad picture of recent findings on the nature of the innovative process. The chapter by Willinger and Zuscovitch is concerned with the new information-intensive production systems. They discuss what is required if firms are to be effective in their R & D, and effective in their use of these new technologies. The chapter by Teece focuses on the firm more narrowly, and explores the question of how conditions of appropriability, and of technological opportunity, affect what it is profitable for firms to do, and the most profitable way for them to pursue various technological options. Kay focuses on the nature of research and development and identifies several key characteristics that influence how it is organized and managed. Coombs is concerned with the new understanding about the reciprocal relationship between technological advance and market structure.

10 The nature of the innovative process*

Giovanni Dosi
Faculty of Statistics, University of Rome, Rome and SPRU, University of Sussex, Brighton

Introduction

The attempt in this book to place technical change at the centre of theory of economic change can draw from a widening empirical evidence on the sources, procedures and microeconomic effects of technical change. Here, I shall try to organise and interpret some of that evidence.

The growing attention to innovation-related phenomena is probably due to various factors, partly internal to the dynamics of the economics discipline and partly related to the increasing empirical perception of the importance of technological factors in competitiveness and growth. For example, an increasing number of industrial case studies has highlighted the importance of technological innovation for industrial competitiveness (for reviews, see Freeman, 1982; Dosi, 1984; Momigliano and Dosi, 1983; OECD, 1984). Moreover, the experience of Japanese competition in international trade and the very rapid productivity gains in Japanese firms have focused attention on a number of features of the Japanese 'national system of innovation' (Aoki, 1986; Altschuler *et al.*, 1985; Freeman, 1987; and Freeman's chapter in this book). Finally, the intuitive perception of the importance of the so-called 'microelectronics revolution', with varying degrees of pessimism on its employment outcomes, has induced a re-appraisal of the possible 'compensation mechanisms', on macroeconomic grounds, between the labour-saving and employment-creating effects of innovation (for critical surveys, see Momigliano, 1985; Freeman and Soete, 1985, 1987; Stoneman, Blattner and Pastré, 1982). Whatever the motivations for the analyses of innovation and technological change, this field of inquiry increasingly highlights the characteristics of a fundamental ingredient of the process of growth and transformation of the economy, which is discussed from different angles in several chapters of this book.

Here, I shall limit my discussion to what I consider some of the 'stylised facts' and fundamental properties associated with the innovative process

*A more detailed and extensive analysis of the origins, nature and effects of innovation—on which this chapter is partly based—can be found in G. Dosi, 'Sources, procedures and microeconomic effects of innovation', SPRU, University of Sussex, Brighton, DRC Discussion Paper. Comments from the participants at the Lewes and Maastricht meetings of the project which led to this book, and in particular those of C. Freeman and R. Nelson, are gratefully acknowledged.

(next section): in doing that, I shall draw from several empirical contributions, over the past decade, on the economics of innovation, including those of Abernathy and Utterback (1975, 1978), Freeman (1982), Klein (1977), Nelson and Winter (1977, 1982), Rosenberg (1976, 1982), Sahal (1979, 1981, 1985), Pavitt (1979, 1984a), von Hippel (1979, 1982), and also some contributions of the author (Dosi, 1982, 1984). Second, I will try to provide an interpretation of technological innovation and its relationship with scientific advances, on the one hand, and market processes, on the other. Finally, I shall argue that such an interpretation of the innovative process is useful to the understanding of inter-industry differences in the modes and degrees of innovativeness, which are further analysed in other chapters of this book.

Some stylised facts on innovation

In an essential sense, innovation concerns the search for, and the discovery, experimentation, development, imitation, and adoption of new products, new production processes and new organisational set-ups. Almost by definition, what is searched for cannot be known with any precision before the activity itself of search and experimentation, so that the technical (and, even more so, commercial) outcomes of innovative efforts can hardly be known *ex ante*. Certainly, whenever innovative activities are undertaken by profit-motivated agents, they must involve also some sort of perception of yet unexploited, technical *and* economic, opportunities. However, such perceptions and beliefs rarely entail any detailed knowledge of what the possible events, states-of-the-world, input combinations, product characteristics will be. Putting it another way, innovation involves a fundamental element of *uncertainty*, which is not simply lack of all the relevant information about the occurrence of known events but, more fundamentally, entails also (a) the existence of techno-economic problems whose solution procedures are unknown (more on it in Nelson and Winter, 1982, and Dosi and Egidi, 1987), and (b) the impossibility of precisely tracing consequences to actions ('. . . if I do this, that will occur . . .', etc.). These uncertainty features of innovative activities are the *first* 'stylised fact'.

Of course, the perception or belief that 'some unexploited opportunity is there' is not always disappointed: the record of technological advances of modern economies, at least since the Industrial Revolution, presents impressive testimony in this respect. In fact, technological innovation has been able to draw, and increasingly so in this century, from novel opportunities stemming from scientific advances (from thermodynamics to biology, electricity, quantum physics, mechanics, etc.). The increasing reliance of major new technological opportunities on advances in scientific knowledge is, in my view, the *second* property of contemporary innovation.

The nature of the *search activities* leading to new products and processes has also changed over the last century: the increasing complexity of research and innovative activities militates in favour of formal organisations (firms' R & D laboratories, government laboratories, universities, etc.) as opposed to individual innovators as the most conducive environment to the production of innovations. Moreover, the formal research activities in the business sector tends to be integrated within more or less integrated manufacturing firms (Mowery, 1983; Teece's and Nelson's chapters in this book). This is the *third* major feature of innovative activities.

However, in addition to the previous point, and in many ways complementary to it, a significant amount of innovations and improvements are originated through 'learning-by-doing' and 'learning-by-using' (Rosenberg, 1976, 1982). That is, people and organisations, primarily firms, can learn how to use/improve/produce things by the very process of doing them, through their 'informal' activities of solving production problems, meeting specific customers' requirements, overcoming various sorts of 'bottlenecks', etc. This is the *fourth* 'stylised fact'.

Fifth, it seems that the patterns of technological change cannot be described as simple and flexible reactions to changes in market conditions: (i) in spite of significant variations with regard to specific innovations, it seems that the directions of technical change are often defined by the state-of-the-art of the technologies already in use; (ii) quite often, it is the nature of technologies themselves that determines the range within which products and processes can adjust to changing economic conditions; and (iii) it is generally the case that the probability of making technological advances in firms, organisations and often countries, is among other things, a function of the technological levels already achieved by them. In other words, technical change is a *cumulative activity*.

How does one interpret these phenomena and link them with the intuitive fact that in market economies particular patterns of innovations find a necessary condition in some sort of (actual and/or expected) economic reward to the innovators? How does one account for the possibility for someone (individuals or firms) being systematically 'better', on technological grounds, than others? What explains the relatively ordered patterns which technical change appears to show and its 'momentum', seemingly propelled by a strong internal logic, despite quite diverse and varying market conditions? I shall now turn to these issues.

Knowledge, opportunities and search: technological paradigms and trajectories

Let me start by observing that the solution of most technological problems (e.g. designing a machine with certain performance characteristics, developing a new chemical compound with certain features, improving the efficiency of a production input, etc.) implies the use of pieces of knowledge of various sorts. Some elements represent widely applicable understanding: it might be direct scientific knowledge or knowledge related to well-known and pervasive applicative principles (e.g. on electricity, mechanics, more recently, informatics, etc.). Some other pieces of knowledge are specific to particular 'ways of doing things', to the experience of the producer, the user, or both.

Moreover, some aspects of this knowledge are well articulated, even written down in considerable detail in manuals and articles and taught in schools. Others are largely tacit, mainly learned through practice and practical examples (of course, 'training' and 'apprenticeship' relate also to this aspect of technology): there are elements of being a 'good engineer', a 'good designer', or even a 'good mathematician' that cannot be entirely transmitted in an explicit algorithmic form.

Finally, some of the knowledge involved in the use and improvement of technologies is open and public: the most obvious examples are scientific and technical publications. However, other aspects are private, either 'implicitly' because they are tacit anyway, or explicitly in the sense that they are protected by secrecy or legal devices such as patents.

All three aspects (universal versus specific, articulated versus tacit, public versus private) are essential in the conceptualisation of 'what is technology'. More precisely, technological advances normally draw on some *sub-set* of the publicly available knowledge, which is shared and improved upon by the community of engineers/applied scientists/designers, etc. However, in the activities aimed at technological innovations, such a shared use of highly *selected* scientific and technological knowledge (related, for example, to selected physical or chemical principles, materials, properties, etc.) is coupled with the use and development of specific and often partly private heuristics and capabilities.

Elsewhere (Dosi, 1982, 1984) as recalled in Chapter 2, I suggest a broad similarity, in terms of definition and procedures although not in objectives or career structures, between 'science' and 'technology'. More precisely, as modern philosophy of science suggests the existence of scientific paradigms (or scientific research programmes), so there are *technological paradigms*. A 'technological paradigm' defines contextually the needs that are meant to be fulfilled, the scientific principles utilised for the task, the material technology to be used. In other words, a technological paradigm can be defined as a 'pattern' for solution of selected techno-economic problems based on highly selected principles derived from the natural sciences. A technological paradigm is both a set of *exemplars*—basic artefacts which are to be developed and improved (a car—of the type we know—an integrated circuit, a lathe, etc., with their particular techno-economic characteristics) and a *set of heuristics*—'Where do we go from here?', 'Where should we search?', 'On what sort of knowledge should we draw?', etc. (consider, for example, general search rules of the kind: 'strive for an increasing miniaturisation of the circuit', 'if an heterocyclic compound worked as a pesticide, fiddle around with the various atomic rings trying to

improve its effectiveness', etc.). Putting it another way, technological paradigms define the technological opportunities for further innovations and some basic procedures on how to exploit them. Thus they also channel the efforts in certain directions rather than others: *a technological trajectory* (Nelson and Winter, 1977, and Dosi, 1982) is the activity of technological progress along the economic and technological trade-offs defined by a paradigm (Gordon and Munson, 1981; Saviotti and Metcalfe, 1984). One can take as fairly evident examples of such paradigms the internal combustion engine, oil-based synthetic chemistry, or microelectronics. A closer look at the patterns of technical change, however, suggests the existence of 'paradigms' and 'trajectories' with different levels of generality, in many industrial sectors.

Freeman and Perez (1986) use the expression 'techno-economic paradigm' to describe those pervasive technologies which influence the behaviour of firms and industries throughout the economic system (see Chapter 3). Note, however, that a 'techno-economic paradigm' (or 'regime') in Freeman–Perez's sense, is a *macro-technological* concept and refers to broad clusters of 'paradigms' in the sense I suggest here: for example, the electronics 'techno-economic paradigm' or 'regime' captures the common characteristics, complementarities and inter-linkages between several 'micro' paradigms—related to semiconductors, computers, industrial automation, etc.

Whatever name is chosen, the concept of 'paradigm' points to interpretations broadly consistent with Rosenberg's 'focusing devices' (Rosenberg, 1976) or Sahal's 'technological guide-posts' (Sahal, 1981, 1985). The crucial hypothesis is that innovative activities are strongly *selective, finalised* in rather precise directions, often *cumulative* activities. This is very different from the concept of technology as information that is generally applicable and easy to reproduce and reuse (Arrow, 1962), and where firms can produce and use innovations mainly by dipping freely into a general 'stock' or 'pool' of technological knowledge. Instead we have firms producing things in ways that are differentiated technically from things in other firms, and making innovations largely on the basis of in-house technology, but with some contribution from other firms, and from public knowledge. Under such circumstances, the search process of industrial firms to improve their technology is *not* likely to be one where they survey the whole stock of technological knowledge before making their technical choices. Given its highly differentiated nature, firms will instead seek to improve and to diversify their technology by searching in zones that enable them to use and to build upon their existing technological base. In other words, technological and organisational changes in each firm are cumulative processes, too. What the firm can hope to do technologically in the future is heavily constrained by what it has been capable of doing in the past. Once the cumulative and firm specific nature of technology is recognised, its development over time ceases to be random, but is constrained to zones closely related technologically to existing activities.

Thus, in general, technological progress proceeds through the development and exploitation of both public elements of knowledge, shared by all actors involved in a certain activity, and private, local, partly tacit, firm-specific, cumulative forms of knowledge.

First, there are certainly 'free-good' elements in technological progress essentially stemming from the free flow of information, readily available publications, etc.

The second aspect of the 'public' characteristics of technology relates to the *untraded interdependences* between sectors, technologies and firms and takes the form of technological complementarities, 'synergies', and flow of stimuli and constraints which do not entirely correspond to commodity flows. All of them represent a structured set of technological externalities which can be a *collective asset* of groups of firms/industries within countries/regions (see, for example, Lundvall, 1984, and Chapter 17 of this book) and/or tend to be internalised within individual companies (see, for example, Teece, 1982, and Chapter 12 of this book; Pavitt, 1984c). In other words, technological bottlenecks and opportunities (Rosenberg, 1976), experiences and skills embodied in people and organisations, capabilities and 'memories' overflowing from one economic activity to another, tend to organise *context conditions* which (i) are country-specific, region-specific or even company-specific; (ii) are a fundamental ingredient in the innovative process; and (iii) as such, determine different incentives/stimuli/constraints to innovation, for any given set of strictly economic signals.

These untraded interdependences and context conditions are, to different degrees, the *unintentional* outcome of decentralised (but irreversible) processes of environmental organisation (one obvious example is 'Silicon Valley') and/or the result of explicit strategies of public and private institutions. (In this sense one can interpret, for example, the strategies of vertical and horizontal integration of electrical oligopolies into microelectronics technologies or the efforts of various governments to create 'science parks', etc.).

To the extent that innovative learning is 'local' and specific in the sense that it is paradigm-bound and occurs along particular trajectories, but is shared—with different competences and degrees of success—by all the economic agents operating on that particular technology, one is likely to observe at the level of whole industries those phenomena of 'dynamic increasing returns' and 'lock-in' into particular technologies discussed in Arthur's chapter (see also Arthur, 1985; David, 1975, 1985).

Conversely, to the extent to which learning is also local and cumulative at the level of *individual firms*, one is likely to observe also firm-specific trajectories, involving the cumulative development and exploitation of internalised (and thus 'private') technological competences, through those strategies discussed in Teece's chapter.

It is important to remark that what has just been said does *not* imply irrelevance of the inducement mechanisms to changes of techniques stemming from the levels and changes in relative prices (in particular, the price of labour to the price of machines (for some recent evidence, cf. Sylos-Labini, 1984) and also to the price of energy and materials, or from changing demand conditions. On the contrary, these factors are likely to be fundamental ones, influencing both the rate and direction of technical progress, but *within the boundaries* defined by the nature of technological paradigms. Moreover, innovation yields new techniques which are likely to be superior to the old ones irrespective of relative prices, either immediately, as often is the case of many microelectronics-based processes (see Soete and Dosi, 1983), or after a learning period (as, for example, in the case of agricultural machinery discussed by David, 1975). If the new techniques had existed before they would also have been adopted at the 'old' relative prices. In other words, technical progress generally exhibits strong *irreversibility features*.

Take the example of microelectronics. As discussed at greater length in Freeman and Soete (1985, 1987), Momigliano (1985), Soete and Dosi (1983), Coriat (1983, 1984), electronics-based production technologies are (i) labour-saving; (ii) fixed-capital saving (i.e. they often induce a fall in the capital/output ratio (for sectoral evidence in the United Kingdom, see Soete and Dosi, 1983); (iii) circulating-capital saving (i.e. the optimisation of production flows allows a fall in the stocks of intermediate inputs per unit of output); (iv) quality-improving (i.e. they increase the accuracy of production processes, allow quality testing, etc.); (v) energy-saving (in so far as the energy use generally is also a function of mechanical movements of the various machineries, the substitution of information processing equipment for electromechanical parts reduces the use of energy). Taking all these characteristics together, it is clear that electronics-based production techniques are generally unequivocally superior to electromechanical ones irrespective of relative prices. That is, the new wage/profit frontiers associated with the new techniques do not intersect for any positive value of the 'old' one (see Dosi, Pavitt and Soete, 1988).

It is important to distinguish between the factors which *induce, stimulate or constrain* technical change from the *outcomes* of the changes themselves. As we analyse in Dosi, Pavitt and Soete (1988), and following the suggestions of Rosenberg (1976), inducement mechanisms may involve a broad set of factors, including:

(a) technological bottlenecks in interrelated activities;
(b) scarcities of critical inputs; or, conversely;
(c) abundance of particular inputs (e.g. energy, raw materials, etc.);
(d) composition, changes and rates of growth of demands;
(e) levels and changes in relative prices (first of all, as mentioned, the relative price of machines to labour);
(f) patterns of industrial conflict.

Where the critical stimuli come from depends on the nature of the technologies and on the economic and institutional context of each country: one can find plenty of evidence on the role of each of these factors. However, irrespective of the immediate triggering factor(s), the patterns of innovation present some remarkable common properties. First, the 'normal' patterns of technological change, to repeat, tend to follow 'trajectories' defined by specific sets of knowledge and expertise. Second, major discontinuities in the patterns of change are associated with changes in technological paradigms (as defined above). Third, irreversibility in the technological advances means also that, using a neo-classical language, the changes *of* the production possibility sets *dominate* over changes *within* any given set. More precisely, at any given time, instead of a well-behaved set we are likely to observe only one (or very few) points corresponding to the best-practice techniques, while, over time, the dominant process of change will imply improvements in these (very few) best-practice techniques (along the 'trajectories'), rather than processes of 'static' inter-factoral substitution.

The conceptualisation of technology and technical change based on 'paradigms', 'guide-posts' or whatever name is chosen, helps also in resolving the long debate in the innovation literature about the relative importance of 'demand pull' (cf. Schmookler, 1966; and, for critical discussions, Mowery and Rosenberg, 1979; Freeman, Clark and Soete, 1982) versus technology push: environment-related factors (such as demand, relative prices, etc.) are instrumental in shaping (a) the rates of technical progress; (b) the precise trajectory of advance, within the (limited) set allowed by any given 'paradigm'; and (c) the selection criteria amongst new potential technological paradigms. However, each body of knowledge, expertise, selected physical and chemical principles, etc. (that is, each paradigm) determines both the opportunities of technical progress and the boundaries within which 'inducement effects' can be exerted by the environment. Moreover, the source of entirely new paradigms is increasingly coming from fundamental advances in science and in the (related) 'general' technologies (e.g. electricity, information-processing, etc.).

So far, I have discussed an interpretation of what I consider fundamental characteristics of the innovative process *in general*. However, at a finer level of analysis, one empirically observes a significant inter-sectoral variety in the rates of technical progress, modes of search, forms of knowledge on which innovation draws. In some areas paradigms are powerful in that they generate rapid sustained technical change. Others are weak, in that they provide relatively little guidance as to where fruitfully to search. Moreover, the fact that a certain kind of technical advance can be achieved cheaply and easily, does not in itself make it profitable for a firm to pursue that advance. I will now discuss these issues.

Opportunities, market conditions and the inter-sectoral differences in innovativeness

On the grounds of the foregoing analysis, the interpretation that I suggest (developing, in particular, on Nelson and Winter, 1982; Freeman, 1982;

Rosenberg, 1976) of the observed differences, over sectors and over time, in the rates and modes by which innovations are generated, diffused and used, traces them back to inter-sectoral and inter-temporal differences in (a) the *opportunities* of innovation that each paradigm entails; (b) the degrees to which firms can obtain economic returns to various kinds of innovation, that is the *degree of appropriability* of innovation; and (c) the *patterns of demand* that firms face.

I have mentioned earlier, among the 'stylised facts', the increasing reliance of major new technological advances upon scientific progress. However, as discussed in detail in Nelson's chapter, only in some technologies and sectors is the link direct and powerful: scientific inputs are, there, an essential part of the momentum of technological advances. In other sectors and technologies the links are much more indirect and may simply relate to the use of science-based equipment and intermediate inputs, or to the generic science-based knowledge acquired by researchers, engineers, etc., during their formal training.

In general, I suggest, the linkages between scientific advances and technological opportunities are likely to be much more direct at the early stage of emergence of new technological paradigms. In these cases, progress in general scientific knowledge yields a widening pool of *potential* technological paradigms. In another work (Dosi, 1984), I analyse the specific mechanisms through which a much smaller set of paradigms are actually developed, economically applied and often become dominant. Here, suffice to say that this process of selection depends, in general, on (a) the nature and the interests of the 'bridging institutions' (Freeman, 1982) between pure research and economic applications; (b) quite often, especially in this century, strictly institutional factors, such as public agencies (the military, space agencies, the health system, etc.); (c) trial-and-error processes of exploration of the new technologies, often associated with 'Schumpeterian' enterpreneurship; (d) the selection criteria of the markets and especially the techno-economic requirements of the users (see Chapter 17 by Lundvall). Certainly, new paradigms become attractive as the cost and difficulty of further progress within existing paradigms increase. However, note that increasing obstacles to progress within a certain paradigm do *not* automatically induce the emergence of new ones; scientific advances are often a necessary condition of their development. Whatever the precise selection mechanisms which produced them, new paradigms reshape the patterns of opportunities of technical progress, in terms of both the *scope* of the innovations and the *ease* with which they are achieved. As examples, the reader may think of the clusters of new technological opportunities associated with electricity, those associated with synthetic oil-based chemistry, or, more recently, microelectronics and bioengineering. Whilst, after a period of intensive development, there might be diminishing returns to innovative efforts within the limits of a *specific* paradigm (the so-called Wolf's Law), new technological paradigms, directly and indirectly—via their effects on 'old' ones—prevent the establishment in general of decreasing returns in the search process for innovations. New paradigms spread their effects well beyond their sector of origin and provide new sources of opportunity, via input/output flows and technological complementarities, to otherwise stagnant activities. The emergence of new paradigms and the diffusion of their effects throughout the economy is possibly the main reason why decreasing returns do not set in throughout the economy: on the contrary, static and dynamic economies of scale are the general rule. Contrary to the most pessimistic expectations of classical economists and contrary also to many contemporary formalisations of problems of allocation of resources in decentralised markets, decreasing returns historically did not emerge even in those activities involving a given and 'natural' factor such as agriculture or mining: mechanisation, chemical fertilisers and pesticides, improved techniques of mineral extraction and purification prevented 'scarcity' from becoming the dominant functional feature of these productive activities. *A fortiori*, this applies to manufacturing.

To summarise: sectors and technologies differ in the easiness and scope of technological advances; these varying technological opportunities depend on the nature of each technological paradigm, on the degrees to which it is able directly to benefit from scientific progress and/or from other new technological breakthroughs, and on its 'maturity'. In turn, paradigm-specific opportunities are a first determinant of the observed inter-sectoral differences in the rates of innovation.

However, for any level of notional opportunities, private, economically motivated agents will invest resources in their exploration only if there is an actual or expected market ultimately willing to pay for it, and if these agents (typically firms) will be able to capture a significant fraction of what the market is willing to pay. In other words, innovative efforts are also a function of *the structure of demand and of the appropriability conditions*: examples of very low innovative efforts by business firms due to lack of appropriability, despite the existence of significant technological opportunities, are discussed in Nelson's chapter.

In general, appropriability conditions differ between industries and between technologies: Levin *et al.* (1984) study the varying empirical relevance as appropriability devices of (i) patents; (ii) secrecy; (iii) lead times; (iv) costs and time required for duplication; (v) learning-curve effects; and (vi) superior sales and service efforts. To these one should add the more obvious forms of appropriation of differential technical efficiency related to scale economies. Of course, the easier it is for firm B to pick up and duplicate the innovative achievements—in terms of product performances or production efficiency—of firm A, the lower the appropriability of innovation. Clearly, with perfect, costless and immediate duplicability no business firm would have any incentive to innovate. Conversely, with very high appropriability only a very little share of the benefits from innovation would spread throughout the economic system in the form of efficiency improvements, learning through imitation and price changes. As it

happens, in contemporary mixed economies one observes, at least within manufacturing, degrees of appropriability which are generally sufficient to provide an incentive to business firms to sustain relatively high rates of technical progress without, however, preventing, sooner or later, imitation, diffusion and distribution of economic benefits to other firms, users and consumers (of course, this is a quite loose proposition since one can hardly define what is the 'sufficient', let alone the 'optimal', degree of appropriability, or how much different innovation would have been under different appropriability regimes, etc.).

In fact, as discussed in Nelson's chapter, Levin *et al.* (1984) find that for most industries 'lead times and learning curve advantages, combined with complementary marketing efforts, appear to be the principal mechanisms of appropriating returns for product innovations' (p. 33). Moreover, there appears to be a quite significant inter-industrial variance in the importance of the various ways of protecting innovations and in the overall degrees of appropriability, with around three-quarters of the industries surveyed by the study claiming the existence of at least one effective means of protecting process innovation and more than 90 per cent of the industries claiming the same regarding product innovations.[1]

If, as suggested, inter-sectoral differences in technological opportunities, appropriability regimes and demand patterns jointly account for the observed inter-sectoral differences in the rates of innovations, these same variables, together with the sector-specific nature of the knowledge on which innovations are based, explain also the sectoral differences in the typical organisational forms of innovative search. For example, some sectors and technologies may mainly rely on 'informal' processes of learning-by-doing and design improvements; others rely heavily on formal search activities undertaken in R & D laboratories; in some sectors innovations are primarily generated by big firms, in others by relatively smaller firms.

Scherer has recently developed an inter-sectoral matrix of the origin and use of R & D in the US economy based on the inter-sectoral generation and use of a large sample of patents (Scherer, 1982). On the grounds of a data base on innovation in the United Kingdom from 1945 to 1979 collected at the Science Policy Research Unit of the University of Sussex, Pavitt (1984a) has developed a sectoral taxonomy of sectors of production and use of innovation. This evidence, and that from Levin *et al.* (1984), seems broadly consistent with the interpretation put forward here of 'why sectors differ in their rates and modes of innovation'.

Pavitt (1984a) identifies from major groups of sectors, namely:

(i) *'Supplier-dominated' sectors* (which include textile, clothing, leather, printing and publishing, wood products). Innovations are mainly process-innovation: innovative opportunities are generally embodied in new varieties of capital equipment and intermediate inputs, originated by firms whose principal activity is outside these sectors themselves. Thus the process of innovation is primarily a process of diffusion of best-practice capital-goods and of innovative intermediate inputs (such as synthetic fibres, etc.). The knowledge base of innovation in these sectors mainly relates to incremental improvements in the equipment produced elsewhere, to its efficient use and to organisational innovations. Appropriability of firm-specific technological capabilities is rather low and firms are typically not very big (with some exceptions in those activities which present economies of scale in production or marketing such as textiles and clothing).

(ii) *'Scale-intensive' sectors.* Innovation relates to both processes and products; production activities generally involve mastering complex systems (and, often, manufacturing complex products); economies of scale of various sorts (in production and/or design, R & D, etc.) are significant; various appropriability devices operate (e.g. lead times, product complexity, etc.); firms tend to be big, produce a relatively high proportion of their own process technology, devote a relatively high proportion of their own resources to innovation, and tend to integrate vertically into the manufacturing of some of their own equipment. This group includes transport equipment, some electric consumer durables, metal manufacturing, food products, parts of the chemical industry, glass and cement. Moreover, within this group one can make a finer taxonomic distinction, according to the nature of the production process, between (a) assembly-based industries (generally characterised by Taylorist/Fordist automation, such as cars, electrical consumer durables, etc.) and (b) continuous process industries (cement, several food products, etc.).

(iii) *'Specialised suppliers'.* Innovative activities relate primarily to product innovations which enter other sectors as capital inputs. Firms tend to be relatively small, operate in close contact with their users and embody a specialised knowledge in design and equipment-building. Typically, this group includes mechanical and instruments engineering. Opportunities are generally high and are often exploited through 'informal' activities of design improvements, introductions of new components, etc. Appropriability is based to a good extent on partly tacit and cumulative skills.

(iv) *'Science-based' sectors.* This group includes the electronics industries and most of the chemical industries. Innovation is often directly linked to new technological paradigms made possible by scientific advances; technological opportunity is very high; appropriability mechanisms range from patents (especially in chemicals and drugs) to lead times and learning curves (especially in electronics); innovative activities are formalised in R & D laboratories; a high proportion of their product innovation enters a wide number of sectors as capital or intermediate inputs; firms tend to be big (with the exception of new 'Schumpeterian' ventures and highly specialised producers.

Admittedly, the empirical evidence on the sectoral patterns and characteristics of the innovation is far from complete. However, my conjecture is that these empirical patterns can be interpreted by means of a few fundamental variables derivable from the relatively general conceptualisation of the process of innovation, outlined above.

Some conclusions

In this chapter, I have tried to analyse some general characteristics of the process leading to the search and economic exploitation of technological innovations in contemporary mixed economies and to apply such a framework to the interpretation of the evidence stemming from a growing number of empirical studies.

The innovative process—it has been argued—entails an intrinsically uncertain activity of search and problem-solving based upon varying combinations of public and private (people-specific or firm-specific) knowledge, general scientific principles and rather idiosynchratic experience, well-articulated procedures and rather tacit competences. I have called a technological paradigm each specific body of knowledge which guides these search and development activities, grows out of the trials and errors of individuals and firms, and is often shared by the entire community of technological and economic actors as the basis upon which one looks for improvements in process efficiency and product performances. Moreover, each paradigm implies different opportunities for innovation, defined in terms of (1) the 'ease' with which technological advances, however defined, can be achieved; (2) different possibilities for the innovator to appropriate economic benefits from it in terms of profits, market shares, etc.; and (3) diffeent degrees of cumulativeness of technological advances in terms of dynamic increasing returns to innovative effort and autocorrelated probabilities of innovative success, either at the level of single firms or industries.

On the basis of this analytical framework, I have suggested some broad conjectures on how inter-technological differences in innovative opportunities, appropriability regimes, knowledge bases, modes of search, etc., might explain the observed variety in the rates and forms of organisation of innovation in contemporary economies.

This interpretation of the innovative process—which draws heavily from the works cited in the introduction and throughout the text—has, in my view, also relevant implications at the levels of both theory and historical analysis.

The theoretical approach with which the present analysis of innovation is consistent (indeed, it is perhaps a necessary microeconomic ingredient) is sketched elsewhere in this book (see, in particular, the chapters in Part II and Part IV) and also, of course, in Nelson and Winter (1982). Here, let me just mention a few implications which might help the reader to grasp some of the analytical threads which hold together the contributions to this book.

First, the foregoing survey of the characteristics of technology and innovation implies a fundamental distinction between *information and knowledge*. Certainly, innovative activities imply imperfect and asymmetric information. However, for whatever available information, the problem-solving activity involved in search and discovery is based on competences, 'visions', and heuristics which are a logical precondition to information processing. Thus this view departs also from any theory of production based on a view of technology, based only or primarily on freely available blueprints (more on this in Winter, 1982; Nelson and Winter, 1982; Amendola, 1983; Dosi and Egidi, 1987).

Second, and relatedly, innovation is generally based on a variety of knowledge sources which inevitably include public institutions, firm-specific experiences and other forms of institution-specific accumulation of competences. Thus the institutional analyses of the chapters that follow in this part and elsewhere in the book are essential to the understanding of the 'anatomy' of the capitalist machine for technological change (cf. especially Nelson's chapter).

Third, as discussed in the chapters by Dosi and Orsenigo, Teece, and Kay, any satisfactory theory of the firm must involve also an institutional (and history-based) analysis of how organisational structures affect the accumulation of competences, and the appropriation of specific rent-earning assets (on this see also Williamson, 1985; Rumelt, 1987; Teece, 1982; Kay, 1984; Pavitt, 1984c).

Fourth, innovative opportunities and their economic exploitation co-evolve in ways that are at least partly endogenous to the process of discovery, development and production, so that the system very seldom hits any 'hard constraint' whereby all available opportunities are fully known and thoroughly optimised. On the contrary, one is likely to observe permanently a variety of search efforts, strategies and results. One can permanently expect to observe (a) inter-firm *asymmetries* in production efficiency and product technologies (cf. the chapters by Metcalfe and Dosi–Orsenigo) and (b) at least equally wide asymmetries among countries. As a consequence, inter-national differences in innovative capabilities, as well as inter-sectoral differences in the patterns of technical change, can be considered as parts of the foundation of a rather general theory of international trade, whereby the sources of competitiveness of each country are not in any meaningful sense a 'primary endowment', but the outcome of processes of innovation, learning, imitation and diffusion (see the chapters by Dosi-Soete, Fagerberg and Perez-Soete; Pasinetti, 1981; Freeman, 1987).

Fifth, and finally, the interpretation of innovative processes briefly outlined in this chapter entails, and is strictly complementary with, a representation of a changing economy as an *evolutionary environment* (cf. in particular the chapters by Silverberg, Metcalfe, Dosi–Orsenego, Coombs, Nelson), wherein economic agents continuously try new things, pay for but also learn from their (and others') mistakes, earn quasi-rents and gain market shares from their success, and, ultimately, contribute to the endogenous evolution of their environment.

Note

1. For detailed discussions of appropriability mechanisms, see also Taylor and Silberston (1973), von Hippel (1979, 1980, 1982), and Buer (1982). The relative costs of innovation versus imitation—clearly a good proxy for appropriability—are studied by Levin *et al.* (1984) and Mansfield (1984). A detailed company-level study of patenting strategies is presented in Wyatt (1985) and Wyatt and Bertin (1985).

References

Abernathy, W. J. and Utterback, J. M. (1975), 'A dynamic model of product and process innovation', *Omega*, vol. 3, no. 6.
—— (1978), 'Patterns of industrial innovation', *Technology Review*, 80, June–July, pp. 2–29.
Altschuler, A., Anderson, M., Jones, D. T., Roos, D. and Womack, J. (1985), *The Future of the Automobile*, Cambridge, Mass., MIT Press.
Amendola, M. (1983), 'A change of perspective in the analysis of the production of innovation', *Metroeconomica*, vol. 35, no. 3, pp. 261–74.
Aoki, M. (1986), 'Horizontal vs vertical information structure of the firm', *American Economic Review*, vol. 76, no. 5, pp. 971–83.
Arrow, K. (1962), 'Economic welfare and allocation of resources for invention', in NBER (1962).
Arthur, W. B. (1985), *Competing Techniques and Lock-in by Historical Events: The Dynamics of Allocation Under Increasing Returns*, Stanford, Stanford University, CEPR.
Baker, M. J. (ed.) (1979), *Industrial Innovation*, London, Macmillan.
Buer, P. (1982), *Investigation of Consistent Make or Buy Patterns of Selected Process Machinery in Selected US Manufacturing Industries*, Cambridge, Mass., Sloan School of Management, MIT, Ph.D. dissertation.
Coriat, B. (1983), *La Robotique*, Paris, La Découverte/Maspero.
—— (1984), 'Crise et électronisation de la production: robotisation d'atelier et modèle Fordien d'accumulation du capital', *Critiques de l'Économie Politique*.
David, P. (1975), *Technical Choice, Innovation and Economic Growth*, Cambridge, Cambridge University Press.
—— (1985), 'Cliometrics and QWERTY', *American Economic Review, Papers and Proceedings*.
Dosi, G. (1982), 'Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change', *Research Policy*, vol. 2, no. 3, pp. 147–62.
—— (1984), *Technical Change and Industrial Transformation*, London, Macmillan.

Dosi, G. and Egidi, M. (1987), *Substantive and Procedural Uncertainty: An Exploration of Economic Behaviours in Complex and Changing Environments*, Brighton, SPRU, University of Sussex, DRC Discussion Paper, presented at the Conference on Flexible Automation and New Work Modes, Paris, 1–4 April 1987.

Dosi, G., Pavitt, K. and Soete, L. (1988), *The Economics of Technical Change and International Trade*, Brighton, Wheatsheaf, forthcoming.

Freeman, C. (1982), *The Economics of Innovation*, London, Frances Pinter, 2nd edn.

—— (1987), *Technology Policy and Economic Performance* , London, Frances Pinter.

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

Freeman, C. and Perez, C. (1986), 'The diffusion of technical innovation and changes of technoeconomic paradigm', presented at the Conference on Innovation Diffusion, Venice, 17–21 March 1986.

Freeman, C. and Soete, L. (1985), *Information Technology and Employment: An Assessment*, Brussels, IBM-Europe.

—— (eds) (1987), *Technical Change and Full Employment*, Oxford, Basil Blackwell.

Gardiner, P. (1984), 'Design trajectories for airplanes and automobiles during the last 50 years'; 'Robust and lean designs', in C. Freeman (ed.), *Design, Innovation and Long Cycles in Economic Development*, London, Royal College of Arts (2nd edn. London, Frances Pinter, 1986).

Gordon, T. J. and Munson, T. R. (1981), *Research into Technology Output Measures*, Gladstonebury, Conn., The Future Group.

Griliches, Z. (1984) (ed.), *R & D, Patents and Productivity*, Chicago, NBER/Chicago University Press.

Hippel, E. von (1979), 'A customer active paradigm for industrial product idea generation', in Baker (ed.) (1979).

—— (1980), 'The user's role in industrial innovation', in B. Dean and J. Goldhar (eds), *Management of Research and Innovation*, Amsterdam, North Holland.

—— (1982), 'Appropriability of innovation benefit as a predictor of the source of innovation', *Research Policy*, vol. 2, no. 2, pp. 95–116.

Kay, N. (1984), *The Innovating Firm*, London, Macmillan.

Klein, B. (1977), *Dynamic Competition*, Cambridge, Mass., Harvard University Press.

Levin, R., Cohen, W. M. and Mowery, D. C. (1985), 'R & D appropriability, opportunity and market structure: new evidence on some Schumpeterian hypotheses', *American Economic Review, Papers and Proceedings*, vol. 75, no. 2, pp. 20–4.

Levin, R., Kleverick, A. K., Nelson, R. and Winter, S. (1984), *Survey on R & D Appropriability and Technological Opportunity: Part 1—Appropriability*, New Haven, Yale University, mimeo.

Lundvall, B. A. (1984), 'User/producer, interaction and innovation', paper presented at the TIP Workshop, Stanford, Stanford University (also printed by Aalborg University Press, Denmark).

Malerba, F. (1985), *The Semiconductor Business: The Economics of Rapid Growth and Decline*, Madison, University of Wisconsin Press.

Mansfield, E. (1984), 'R & D and innovation: some empirical findings', in Griliches (ed.) (1984).

Mansfield, E. *et al.* (1971), *Research and Innovation in the Modern Corporation*, New York, Norton.

Metcalfe, J. S. (1985), 'On technological competition', Manchester, Department of Economics, University of Manchester, mimeo.

Momigliano, F. (1985), 'Le tecnologia dell'informazione: effetti economici e politiche pubbliche', in A. Ruberti (ed.), *Tecnologia Domani*, Bari, Laterza-Seat.

Momigliano, F. and Dosi, G. (1983), *Tecnologia e Organizzazione Industriale Internazionale*, Bologna, Il Mulino.

Mowery, D. (1983), 'The relationship between intrafirm and contractual forms of industrial research in American manufacturing, 1900–1940', *Explorations in Economic History*, vol. 20, no. 4, pp. 351–74.

Mowery, D. and Rosenberg, N. (1979), 'The influence of market demand upon innovation: a critical review of some recent empirical studies', *Research Policy*, vol. 8, pp. 102–53.

National Science Foundation (1983), *The Process of Technological Innovation: Reviewing the Literature*, Washington DC, National Science Foundation.

NBER (National Bureau of Economic Research) (1962), *The Rate and Direction of Inventive Activity*, Princeton, Princeton University Press.

Nelson, R. (1980), 'Production sets, technological knowledge and R & D: fragile and overworked constructs for analysis of productivity growth?', *American Economic Review, Papers and Proceedings*.

—— (1981), 'Research on productivity growth and productivity differences: dead ends and new departures', *Journal of Economic Literature*, vol. 19, no. 3, pp. 1029–64.

—— (1982), 'The role of knowledge in R & D efficiency', *Quarterly Journal of Economics*, vol. 96, no. 3, pp. 453–70.

Nelson, R. and Winter, S. (1977), 'In search of a useful theory of innovations', *Research Policy*, vol. 6, no. 1, pp. 36–77.

—— (1982), *An Evolutionary Theory of Economic Change*, Cambridge, Mass., The Belknap Press of Harvard University Press.

OECD (1984), *Committee for Scientific and Technological Policy, Science, Technology and Competitiveness: Analytical Report of the Ad Hoc Group*, Paris, OECD/STP (84) 26.

Pasinetti, L. L. (1981), *Structural Change and Economic Growth*, Cambridge, Cambridge University Press.

Pavitt, K. (1979), 'Technical Innovation and Industrial Development: The New Causality', *Futures*.

Pavitt, K. (1984a), 'Patterns of technical change: towards a taxonomy and a theory', *Research Policy*, vol. 13, no. 6, pp. 343–74.

—— (1984b), 'Chips and "trajectories": how will the semiconductor influence the sources and directions of technical change?', in R. Macleod (ed.) (1986), *Technology and the Human Prospect*, London, Frances Pinter.

—— (1984c), *Technological Innovation and Strategic Management*, Brighton, SPRU, University of Sussex, mimeo.

Perez, C. (1985), 'Microelectronics, long waves and the world structural change: new perspectives for developing countries', *World Development*, vol. 13, no. 3, pp. 441–63.

Rosenberg, N. (1976), *Perspectives on Technology*, Cambridge, Cambridge University Press.
—— (1982), *Inside the Black Box*, Cambridge, Cambridge University Press.
Rothwell, R. *et al.* (1974), 'SAPPHO Updated. Project SAPPHO, Phase 2', *Research Policy*, vol. 3, no. 3, pp. 258–91.
Rothwell, R. and Gardiner, P. (1984), 'The role of design in product and process change', *Design Studies*, vol. 4, no. 3, pp. 161–70.
Rumelt, R. (1987), 'Theory, strategy and enterpreneurship', in D. Teece (ed.) (1987).
Sahal, D. (1979), *Recent Advances in the Theory of Technological Change*, Berlin, International Institute of Management.
—— (1981), *Patterns of Technological Innovation*, New York, Addison Wesley.
—— (1985), 'Technology guide-posts and innovation avenues', *Research Policy*, vol. 14, no. 2, pp. 61–82.
Salter, W. (1960), *Productivity and Technical Change*, Cambridge, Cambridge University Press.
Saviotti, P. and Metcalfe, J. S. (1984), 'A theoretical approach to the construction of technology output indicators', *Research Policy*, vol. 13, no. 3, pp. 141–52.
Scherer, E. (1982), 'Inter-industry technology flows in the US', *Research Policy*, vol. 11, no. 4, pp. 227–46.
Schmookler, J. (1966), *Invention and Economic Growth*, Cambridge, Mass., Harvard University Press.
Soete, L. (1979), 'Firm size and innovative activity: the evidence reconsidered', *European Economic Review*, vol. 12, pp. 319–40.
Soete, L. and Dosi, G. (1983), *Technology and Employment in the Electronics Industry*, London, Frances Pinter.
Stoneman, P., Blattner, N. and Pastre, O. (1982), 'Major findings and policy responses to the impact of information technology on productivity and employment', *Microelectronics, Robotics and Jobs*, Paris, ICCP, OECD.
Sylos-Labini, P. (1984), *Le Forze dello Sviluppo e del Declino*, Baris, Laterza (English translation: *The Forces of Development and Decline*, Cambridge, Mass., MIT Press, 1985.
Taylor, C. and Silberston, A. (1973), *The Economic Impact of the Patent System*, Cambridge, Cambridge University Press.
Teece, D. (1982), 'Toward an economic theory of the multi-product firms', *Journal of Economic Behaviour and Organisation*, vol. 3, no. 1, pp. 39–64.
—— (ed.) (1987), *Strategy and Organisation for Industrial Innovation and Renewal*, New York, Ballinger.
Williamson, O. (1985), *The Economic Institutions of Capitalism*, New York, Free Press.
Winter, S. (1982), 'An essay on the theory of production', in S. H. Hymans (ed.), *Economics and the World Around It*, Ann Arbor, University of Michigan Press.
—— (1984), 'Schumpeterian competition in alternative technological regimes', *Journal of Economic Behaviour and Organisation*, vol. 5, no. 3–4, pp. 287–320.
Wyatt, S. with Bertin, G. and Pavitt, K. (1985), 'Patents and multinational corporations: results from questionnaires', *World Patent Information*, vol. 7, no. 3, pp. 196–212.
Wyatt, S. and Bertin, G. (1985), *The Role of Patents in Multinational Corporations Strategies for Growth*, Paris, AREPIT.

11 Towards the economics of information-intensive production systems: the case of advanced materials*

Marc Willinger and Ehud Zuscovitch
Bureau d'Economie Théorique et Appliquée, Université Louis Pasteur, Strasbourg

Introduction

The ongoing shift from one technological system to another is not just a matter of setting up a new configuration of advanced technologies. A qualitative change in the nature of economic activity is under way and the industrial system already looks radically different compared with the post-war system (see Freeman and Perez, and Boyer in this volume). One of the main features of the new technical system appears to be its 'information intensity' meaning, firstly, that the system is capable of creating and dealing with increasing amounts of information, and, secondly, that it becomes more and more able to adjust itself to a growing and changing variety of signals generated by the economic environment. In this chapter we shall study the implications of these features for the behaviour of firms. We shall show how the management of technology changes under Information-Intensive Production Systems (IIPS). Efforts to achieve scale economies partly shift from standardisation of R & D and production technologies to the coordination of specialities with a view to dominating increasing complexity.

In order to grasp this qualitative change, we shall analyse in the first section, 'likely technological trends', the management of technological options and the accumulation of technological knowledge during the development of a usual technological paradigm. The emphasis on information, learning and the kind of technical properties that one looks for during the diffusion process will prepare the ground for the question of dealing with growing information intensity.

We shall try to tackle this issue in the subsequent section, 'Permanent variety and changes in firms' strategies', after a proper characterisation of IIPS. With less standardisation, the selection process operates less

*The research whose results are reported here was performed under the FAST project (Forecasting and Assessment in Science and Technology) of the EEC DG XII, which we gratefully acknowledge for support. See '*New Advanced Materials*', P. Cohendet, M.J. Ledoux & E. Zuscovitch Eds, Springer-Verlag 1988. We have benefited from discussions with participants in the IFIAS meeting at Maastricht in May 1987. The editorial work of R. Nelson and G. Dosi helped much towards the reshaping of this chapter into its present form; as did the remarks by C. Freeman and N. Kay.

vigorously so that technological options become less and less mutually exclusive. Two major problems arise. The first is that no firm is singly capable of mastering the growing informational requirements (with regard to both technological knowledge and markets information). Indeed, various types of cooperation have been developing rapidly for some time now. The second is that with fewer scale economies of the standard type one needs another; otherwise, the whole system is not viable. Informational scale economies, linked to flexible integration of decisions and routines, may offer an alternative.

The area in which we have done the research leading to this chapter is that of advanced materials. The choice was not random. On the one hand, substitution and diffusion mechanisms can be very nicely illustrated with materials because each of them has a large spectrum of product applications. On the other hand, new materials are an essential component of the new technological system. While information technologies generate flexibility, advanced materials support the induced product variety. One development, we suggest, reinforces the other.

Likely technological trends

Increasing variety of products and their management, or what we call IIPS should be carefully delimited because increasing variety is also a typical property of transition situations. Actually, whenever the economy goes through a structural change caused by the modification of the technological basis upon which it is established, numerous new technical options are tested before the selection process, partly through lock-in mechanisms and partly by scaling-up and cost reduction, establishes the new standard that will become the dominating technological paradigm. In transition periods, as uncertainty rises, agents reconsider the composition of their R & D activities so as to be, if possible, at the right place when important choices in development trends are to be made. This adjustment attitude in itself tends to increase variety because most firms combine passive behaviour of 'just keeping oneself informed' with active R & D behaviour generating new options. Using diversification to reduce uncertainty was one of the driving forces behind the rapid penetration of information technologies. Yet the latter, once introduced, increase variety by their own nature, and push the system towards growing information intensity. It is, then, necessary to distinguish between the set of arguments relative to *transitional variety*, which disappears once the paradigm is stabilised, and the one that develops out of the very nature of IIPS.

Therefore we shall study here the emergence and management of variety in transitory situations, that is, at the pre-paradigmatic stage (see next sub-section). We shall then analyse the nature of the substitution process which takes place through the change from one paradigm to another. This will highlight the technology management problem as a function of the knowledge requirements before and after the stabilisation of the new paradigm.

Technological learning and technological options

Let us consider the example of new materials. Their development generates a growing variety in the material industries, thereby increasing the number of technological options. How these options are gradually revealed and defined depends on the particular way new materials are introduced in an already existing technological framework. As we shall see, this is done in a piece-by-piece substitution process in which learning effects lead to a global reconception of complex products such as cars or aeroplanes. Firms try to manage the overall process by a broadening of their range of activities, thus generating technological variety in order to prepare (and to be prepared for) selection. They often aim at *horizontal integration* within a whole set of materials. Synergy effects of being familiar with the relevant markets, as well as a knowledge of neighbouring production processes, are more likely to be profitable than purely random diversification (see Teece's chapter). This, for example, partly explains why the chemical industry goes into ceramics activities, or why some steel producers take part in the plastics industry.

Maintaining variety is also part of the strategies through which firms try to manage substitution phenomena. Some firms specifically invest in new materials in order to slow down their diffusion. This behaviour may seem paradoxical, but it can be illustrated by an example from the aluminium industry. Composite materials in the aerospace applications are a major threat to aluminium producers, who are the traditional suppliers of the market. To match the challenge, aluminium producers invest in composite materials (especially carbon fibres), while simultaneously developing new and improved aluminium-lithium alloys (Al-Li).[1] As a result they control and reduce the range of substitution possibilities between composites and aluminium alloys. However, aluminium producers are aware of the fact that, in the long run, major substitutions may well occur just the same. Their investment in carbon fibres can also be viewed as a long-term strategy to enter the composite market. In the short run, however, their diversification behaviour is directed towards valorization of both past and current research on new alloys (Al-Li) and appears as a means to secure profitability of existing capital and thus to recover past investment.

Nevertheless, generally speaking, investment in new materials and new technologies is made in order to acquire knowledge about them. If firms expect that in the near future most of today's technical problems will be solved, there is an incentive to gather production knowledge today. There is an *economic advantage in being ready* at the moment when the market selection process reveals its outcome. To illustrate this principle let us take the case of newly developed ceramics. Among the wide field of potential applications for these materials, those of particular interest are the thermo-mechanical applications (engines, turbines, etc.) in which ceramics may progressively replace metallic alloys. In the short run, however, most expectations conclude that metallic alloys will keep their dominant position. This is mainly due to the fact that, facing a risk of substitution, the metal industry develops improved alloys with better properties (recall Rosenberg's example of sailing ships).[2] But the industrial control of ceramics production is an immediate challenge and, indeed, Japanese firms have already started industrial production of ceramics for thermo-mechanical applications, in spite of the fact that today the market is still very narrow. This production is not actually intended to be sold as such, but is primarily made in order to learn how to produce ceramics. It is merely a kind of *technological simulation*. Scientific and technical knowledge-acquisition is one thing, but in view of competition the crucial stake is production-knowledge.

It is important to note that such strategies act in favour of the emergence of new materials and play an important role in the rate and direction of the innovating process (it is a kind of self-fulfilling expectation in technical progress).

Diffusion patterns: complementarities, substitution and the redesign of systems

At any one time the field for material applications can be divided into two types of area. There are areas in which only one specific material can resolve the relevant technical problem. There are others in which a number of materials compete in order to fulfil the same function. In an area where only one material imposes itself there is obviously no economic problem of inter-technological rivalry, which, on the contrary, is important in the latter cases and operates also through the relative prices of resources employed in the manufacturing of various materials.

The simplistic description that we have just presented helps one to grasp the dynamics of the development of a new material. The technical needs which stimulate the emergence of a new material are clearly those for which the existing materials, produced in well-established technologies, give neither technically nor economically satisfying solutions. It is in the periphery of the field of existing technical paradigms that solutions are not very effective and are consequently expensive (these are the zones of diminishing returns for standard materials). Facing this stimulus, some firms begin to develop new materials which can respond to technical challenges which are either insoluble or not economically viable through the usual means. The cost of development of these new materials is important but, in spite of that, their advantage is that they could potentially compete with the traditional materials in some special applications.

The R & D carried out at this stage of emergence of a new material is of a particular nature. Given the high cost of the new elements, an incremental procedure is generally adopted. The designer will typically insert a new piece made of new materials in *existing technical devices* and the new item will have to match up to the old ones. For example, some parts such as piston heads will use new ceramics in an otherwise metallic engine. Another example is the use of elements made of composite materials in an aircraft within the existing aluminium design. This piece-by-piece substitution process implies that the R & D will have to handle compatibility properties, and *complementarity* links are created between old and new materials.

Through this incremental process learning occurs. As scientific and technical problems are progressively solved, new processing techniques are elaborated and thus knowledge and know-how are created, skill networks and industrial standards emerge, scale economies appear, and cost reduction along with it. At a certain stage, the diffusion process goes through a *threshold*. Technical devices are completely reconceived as a function of the properties of the new compounds: consequently, the R & D will look for *substitution* properties in order to be able to compete with other materials, even in the fields which were hitherto exclusively reserved for them. This 'materials war' involves phenomena of irreversibility (in this area, as well as more generally in several other technologies, see David, 1975; Rosenberg, 1982). If at a given moment, in order to satisfy a particular need, a specific material imposes itself with some scale it will tend to exclude others by the mere fact that R & D budgets are limited and options that are not actually in practice are not improved. Consequently, they cease their evolution and, very often, become impossible (or too expensive) to reactivate (Zuscovitch, 1986; and Arthur's chapter).

A technical system such as an aeroplane is a system of interdependence of technical constraints favouring the incrementalism of innovations. The variability in the innovative options on any one component depends on the constraints coming from the features of other components. The place which was formerly occupied with a metal piece, and in which one wishes to insert another made of composite material, was defined as a function of the whole (metallic) system. Now, there is no reason why a composite should behave like a metal and so using it within the system will make the users discover the new properties such material exhibits. To take full advantage of these properties one would have to *redesign*:

> It is not just simply a case of substituting a composite for a metal, but rather of completely redesigning each element, profiting from the (directional or multi-directional) characteristics of the composites used, and very quickly we realize that the rational use of composites brings upon the reconsideration of the very way one conceives and manufactures structures.[3]

Examples of such reconceptions can, of course, be found not only in aeronautics (the rotor boss of the SA 365 c helicopter), but also in the car industry (Bertin's suspension), and in the railways (the bogie elaborated by MBB). In most cases these reconceptions are characterised by a new integration of pieces and a simplification of the system.

This simplification is no trivial operation and, in general, the reconception of technical systems such as an aircraft, in order to take into account the properties of a new material, is a complex operation. The computerisation of conception methods, thanks to CAD, has made a qualitative jump possible in the field. In spite of this, however, the reconception of systems remains above all a function of the constitution of new bodies of knowledge. The latter is only possible through a cumulative process which can only start precisely by a 'piece-by-piece' substitution. The emergence of a threshold effect working in favour of the complete reconception of a system is only realised after the exhaustion of gains from an incremental type of logic. It is often at the moment when expected marginal returns from the incremental innovation are negligible that the technical knowledge, accumulated at the time of the 'piece-by-piece' substitution process, reaches the critical mass necessary for the global reconception of the system.

This critical amount of knowledge is reached through the growth of the number of links, which one has to master in order to integrate pieces made of one material into systems originally created for different ones. The complexity of both system and knowledge increases to the point where a simplification of the system is called for (and allowed). During the 'step-by-step' substitution periods, firms search and, consciously or not, relax the technical constraints of the 'old' optimal solution. As the process goes on, knowledge concerning advantageous reconception opportunities is gained. When the field of such opportunities is sufficiently opened, a qualitative jump becomes possible on the technical level as well as on that of economic performance.

This reinterpretation of the diffusion process in terms of technological learning and the management of technological options suggests that an important problem may arise. Under IIPS, where increasing variety and the capacity *permanently* to maintain many options are decisive, the usual capital-increasing scale economics are likely to operate less effectively. This will consequently alter the nature of economic activities within and among firms. We shall study this problem in depth in the next section.

Permanent variety and changes in firms' strategies

The learning process described in the first section involved a selection mechanism which reduced the number of technological options. The latter were actually mutually exclusive, at least after some stage of development. It follows that the information was intensive (in knowledge and know-how variety) only before selection. Growing flexibility and induced product variety modifies this basic tendency. We shall first try to characterize the technological and market knowledge requirements in IIPS. With many coexisting alternatives, diffusion curves should become smaller and flatter and the problem of economic viability will immediately arise. The issue can be divided into two separate questions. The first is the question of the viability of the resource allocation process in IIPS. The second concerns

the dynamic aspect and the possible existence of a sort of 'increasing returns to information', somewhat equivalent to the capital-scale economies in the mass production system.

Characteristics of IIPS

The keystone of the technical system which is still dominant today is *the standardisation of production*. It relies upon a process of technical and organizational changes in order to scale up operations and to satisfy a growing demand. However in the post-war period, the quick growth and the constant increase of welfare in industrial countries had its price. The logic of standardisation has often implied a very limited flow from science to technology and production: only a tiny part of science-based knowledge could be implemented. Moreover, the satisfaction of consumer needs occurred through standardised products so that consumers could benefit from the advantages of mass production.

The characteristics of the materials industries are very typical of that industrial context. The post-war technical system was built up on quite a large set of materials such as various metals, plastics, concrete, glass, wood, etc. The production of these materials was, nevertheless, highly standardised, each one being produced in large quantities and representing an essential component of the general mass-production nature of the economic system. In actual fact, the larger part of manufacturing as a whole was based upon, and often structured around, a limited number of materials (e.g. metal products). While suitable for a large spectrum of applications, that set-up nevertheless restricted the scope for product variations and the industry's capacity for adjustment. Specialised materials of course already existed in order to match specific needs and applications, but more often than not they were primarily produced as standard commodities and then subsequently adapted and 'functionalised'.

Since the late 1960s, production technologies have undergone a qualitative change. The change was partly induced by endogenous technological developments and partly by more general economic factors. First, during the period of fast growth per capita income rose constantly: a richer society demands more variety and Western countries exhibit the rich man's problems. Second, following the oil crisis, the deep uncertainty about both energy prices and the nature of the ongoing technical change has considerably increased the need for flexibility. Yet without the technological push of both information technology and advanced materials the working rules of the economic system would not change that much.

In effect, the origin of the changes in the production system goes back to the early programmable automation. From the start, numerically controlled machines showed that, along with the usual benefits of automation, smaller series also became competitive and allowed a larger product variety than before. In the late 1970s and even more so in the 1980s, a rapid diffusion of micro-computer technologies in production took place in the forms of computer-aided design, computer-aided manufacturing, robotics, artificial intelligence applications and flexible manufacturing systems. In almost all industrial sectors this embodiment of 'intelligence' in production tools considerably lowers the overall set-up costs and thus enables the producer to switch more often from one product to another. Increasingly it seems that efficiency and variety are no longer rival objectives. In a growing part of the industry, from clothing to leather, plastics processing, metal products, etc., it becomes increasingly profitable to produce small batches with a wide spectrum of combinations of properties.

In the materials industry a parallel evolution supplies the necessary condition for the new system. When the variety of products increases, materials have to adjust and offer 'tailor-made' solutions which exhibit the required properties for each particular application, optimising both users' and producers' constraints. While some adaptation in the materials used has always been necessary, it was usual practice to choose the material whose features best suited the most important technical requirement and then took the other features of that material as constraints upon design and processing. Now the materials choice itself has become increasingly an endogenous design variable, that is, a dimension subject to programming. In turn, this has become possible due to both a growing understanding of the microscopic properties of matter and the development of new processing technologies. In other words, new materials loosen one of the tightest constraints on the evolution of the global technical system.

As a consequence, in the new economic system which is gradually emerging, the firm plays the role of *coordination of inputs and outputs properties through network management*. The firm is increasingly trying to match the variety of properties demanded by the user with its own technological 'data base'. Such a task requires permanent research into the characteristics demanded by the user, on the one hand, and the development of the scientific and technical knowledge that is likely to be used in fulfilling this demand, on the other. The definition of the product can no longer be made without the active cooperation of the user, who must reveal his needs and preferences and participate in the definition of the necessary technical solution.[4]

The informational viability of the firm

In an Information-Intensive Production System (IIPS) firms have to face increasing information costs. A growing amount of resources and effort is required for informational activities, namely (a) gathering information on specialised and changing micro-markets, and (b) searching for new technical solutions adapted to the specific needs revealed by these micro-markets. As a result firms have to cope with higher transaction costs to implement refined market searching and with higher R & D expenditures. Hence, one arrives at a basic viability question: how do the required resources for developing such activities become available?

One part of the possible answer comes from the consumer's willingness to pay: tailor-made products are higher priced than standardised products.

Indeed, lower price-elasticity of demand and quasi-monopoly power in differentiated-product industries lead to higher selling returns than in competitive industries. However, whether these higher returns will fall short of or exceed the increased informational costs ultimately depends on *the size of each 'micro' market*. In general, for any given micro-market size, and level of the quasi-rent on the specific product-service package, the basic question is whether there exists a reallocation of resources within the firm which is economically viable, i.e. which can at least offset the increased informational costs.[5]

Of course, the question of viability of a *variety-based, information-intensive production system* is directly linked to the question of its permanence. When such viable reallocations do not exist, variety can only appear as a transitory phenomenon that will vanish after the market selection process has elicited some new standards. If, for example, small series are less efficient within a given market, despite potential variety in demand, such a selection process would push towards scale industries and standardisation of products.[6] The post-war technical system should be considered as a result of such a process which favoured large-scale production in most industries. We shall try to tackle the question of viability of IIPS by discussing, first, increasing transaction costs and, second, growing R & D expenditure.

(a) Transaction costs of micro-markets and economies of scope. Increasing transaction costs arise mainly because of specialised demands which, in a dynamic context, are frequently changing. In a variety system such costs are no longer entirely fixed but are mainly variable, and arise because of the time and costs required to locate, define and satisfy more specialised needs. In the post-war technical system, transaction costs were a relatively small proportion of the total cost of the firm. In addition, they could be shared among a large number of identical products. Economic viability was granted through scale economies obtained with specific capital goods. In a variety-based system, however, large series of identical products are replaced by small batches of a broad range of different products. With small series, scale economies are severely reduced if specific capital goods are employed. Unit costs will increase as well as transaction costs. However, new equipment such as CAD/CAM, robotics and flexible manufacturing systems (FMS) exhibit a kind of advantage that could be compared to scale economies. For example, robot production lines incorporate enough flexibility to handle a variety of different products. More generally, by sharing a given set of inputs *economies of scope*[7] may be obtained. Economies of scope arise mainly when the 'horizontal constraint', i.e. specialisation in the use of capital equipment, is removed. This is indeed the case with FMSs which are not designed for specific tasks. Their scope of operation is quite large and such equipment easily adapts to new production lines simply by modification of the software. As a consequence variety and efficiency no longer appear as rival objectives. Joint

production costs obtained by sharing the same capital inputs are lower than the sum of specific production costs. To put it another way, product-specific economies of scale are smaller than economies of scope. Such a property, called 'transray convexity'[8] of the firm's cost function, means that 'as a firm changes the composition of output while holding fixed the level of some aggregate measure of output, costs will be lower for diverse rather than specialized output mixes' (Bailey and Friedlaender, 1982).

Traditional economies of scale still appear when aggregate output increases, as the result of physical indivisibilities of capital goods. But if full capacity is not attained with given equipment and for some product mix, it will be efficient to fill the gap by adding a new product to the mix. Now if economies of scope are high enough, they compensate, at least to some extent, the increasing costs resulting from information-gathering activities. Thus one can argue that economic viability is likely, at least for certain degrees of variety. However, economic viability does not depend only on the general properties of a technological system (in our case, the IIPS), but also on the firm-specific, and path-dependent, abilities of individual firms (cf. in this volume the chapters by Teece and Arthur). The viability of IIPS for each firm crucially depends on its success in adjusting its behavioural routines—especially with regard to R & D activities—and in acquiring the right kind of knowledge. We shall now discuss these fundamental requirements.

(b) R & D management in IIPS and organisational flexibility. In a standardised production system, R & D expenditures are mainly fixed costs for a given product and research costs are distributed among a great number of identical products. In addition, informational economies of scope within R & D activities occur and indeed many improvements are readily applied to products and technologies belonging to the same family. Output of R & D activities are shared inputs for production activities. As an example, take the case of the French nuclear power industry. All power stations were built according to the same standards. As a result, any improvement made in one station, or knowledge acquired within a given process, was automatically applied or used in others.

We suggest that such advantages no longer hold (or are less important) within an IIPS. First, R & D costs are distributed only over quite a narrow set of identical products. Second, improvements made in a given process are not easily applied to other processes. Because of greater variability of R & D expenditures firms are faced with a new cost constraint, and the question is once again one of viability in dealing with these increasing R & D costs.

At a first glance, a high variety of R & D projects under these circumstances seem hardly viable. Indeed, this is likely to be the case if R & D is managed the same way as in the 'standardised' production system. However, one may conceive of different organisational developments capable of handling an increased output variety and input specificity. One

such development involves a recombination of standardised intermediate consumption in order to obtain output variety. As a result, final products will be of variable design but will incorporate mainly standardised components. Only some characteristics of the products will be more variable at the expense of higher standardisation of other characteristics. Such a possible evolution was pointed out by Cohendet and Llerena (1987), and is well exemplified by some material industries, such as silicium production, for example. In this case new technologies and materials would only affect the distribution of variety and standardisation among industries without affecting the average degree of variety within the technical system.

However, increasing variety in some industries does not necessarily imply higher standardisation in others. Another development with full exploitation of IIPS is nevertheless possible, subject to two conditions. The first is the formation of technological partnerships; the second is the capacity to internalise efficiently external information. These two points will be developed.

First, under these circumstances, R & D viability rests upon a radical change in its organisation. If variety is an economic objective, concentrated research activities are less efficient because the results of each R & D project are distributed over a small set of products each with relatively short production runs. This is the main reason that induces firms to collaborate with others instead of integrating new activities or doing 'in-house' research, illustrating Teece's observations on the limits of integration (Chapter 12). Thus firms will have to develop more cooperative research, especially with other firms possessing complementary knowledge and skills. This is indeed the case with composite materials. Development of composite products involves quite different bodies of technical knowledge. Most R & D activities are undertaken by teams organised through partnership of different firms. As a result firms also try to develop their ability to organise such cooperative research activities, obviously involving, for these tasks, resources and organisational effort. In turn, this has some major implications concerning the organisation of the firm, and especially its communication system, i.e. its internal and external system of interactions.

A firm will try to set up an optimal communication network internally in order to save on communication costs within the firm and to take into account the economic environment. Once such a communication network is installed, set-up costs are not easily recovered because such networks are highly firm-specific, depending in particular on the firm's 'culture'. There is then a natural tendency towards specialisation of networks in order to achieve efficiency with respect to a given environment. On the other hand, an already existing network can generate economies of scope because it can be used for a great deal of different messages. But as Arrow noted, optimality of the firm's communication system is no longer granted if external conditions change:

Eventually the communication system may be very inefficient at handling signals, and the firm may vanish or undergo a major reorganization. To put it in another way, the firm's organization is designed to meet a more or less wide variety of possible signals. The wider the range planned for, the greater is the flexibility of the firm in meeting the unforeseen (that is what flexibility means) but the less efficient it is in meeting a narrower range of possibilities . . . [Arrow, 1973]

A similar point is made in relation to user-producer communications by Lundvall in Chapter 17. He relates the rigidity of the communication system to the efficiency of innovations. A given set of communication systems may eventually lead to 'unsatisfactory innovations' when consumer needs change. However, firms must not only be flexible in their handling of the information from the environment. They must also be able to draw from the appropriate technical knowledge and skills distributed in the environment. Let us call such threads of technological flows and interdependences (cf. the chapters by Lundvall and Dosi) the *skill network* of an environment.

In IIPS, in order to secure viability, firms have to create new relations with the outside skill network in order to produce new and adapted solutions for evolving needs. This means that firms have to develop organisational flexibility, i.e. the capacity to generate and organise new relations within their environment and especially with the skill network.[9] Such organisational flexibility seems to be a main issue for competition in IIPS.

The firm in IIPS and the 'internalisation of the environment'

Higher product variety and rising information intensity imply that firms have to manage growing complexity in information-processing and problem-solving (cf. Dosi–Orsenigo in this volume). This can only be dealt with by *simplifying internal procedures*, and by developing *coordination skills*. Such integration is capable of liberating the firm's human resources required to deal with external information.

(a) Informational scale economies through algorithmisation. The capacity permanently to process new information, which also amplifies the strategic reach of the firm, implies the capacity to transform an increasing number of choices and assessments into routinised procedures. There must be a continous transformation of problems that belong to the field of decision-making into that of standardised responses. It is only in so far as such a transformation operates with a certain ease and autonomy that the full IIPS is viable. We call such a process the 'algorithmisation' of information-processing, decision-making and organisational coordination. We conjecture that, on the grounds of the new technologies, such a process (a) leads to the economic viability of IIPS, and (b) embodies a momentum of its own, like a 'trajectory', in the terminology of Dosi's chapter, towards higher levels of efficiency with flexibility. This trajectory is, loosely speaking, the equivalent in IIPS of the trajectory toward mechanisation, automation and economies of scale in the earlier 'standardised' production system.

It is necessary, in a way, to prove whether the new functioning principle based on the processing of information is capable of *generating a surplus*. During industrialisation the principle of surplus creation, and hence of economic development, relied on a chain reaction that linked *standardised* organisation of work, dispossession of individual qualifications by automisation and opening of bigger and bigger markets. The static expression given in economics textbook to this process is increasing returns to scale. For the new information-intensive production regime to be viable, a new chain reaction must take place without being conditioned by standardisation in the usual sense.

One is already able to see some steps in the 'chain-reaction' which led the evolution of information technologies. With the diffusion of central computers in the 1960s, firms were able, by batch processing, to improve the efficiency of their routine functions (accounting, stock management, billing, etc.). At the end of the 1970s and at the beginning of the 1980s, this was followed by the development of more decentralised computer systems and the integration of computerisation of production, via CAD, robotics and, finally, the flexible workshop. The latter step brings about a change, we suggest, in the general logic of production organisation, capable of dealing more efficiently with frequent changes. However, this second stage has been largely conditioned by the previous ones. Indeed, the capacity to move from one production sequence to another requires the optimisation of each. This optimisation is itself a result of the accumulation of knowledge and control of the processes generated by the previous stages of informatisation. For example, when in the cost-accounting area the use of computers creates better knowledge of, say, factors that influence material consumption, new instructions to production management are given. In turn, codified instructions mean that, in areas where decisions had previously to be made, a routine procedure is implemented. The rate of 'algorithmisation' consequently increases in the workshop, bringing it closer to the level at which it can integrate its own computers. It is the control of the production organisation via computerisation that allows codified changes in products and processes and hence makes a wider variety of management possible.

The computerisation of standardised functions generates a 'surplus of information' on less well-organised functions. As a consequence this allows a more accurate determination of the behavioural norms related to these functions, whereas before their decision-making was more uncertain. These more 'discretionary' functions become in turn standardised, and thus more apt to absorb information technologies in their own right. Apparently what appears here is a chain-reaction process in which each computerised stage prepares the necessary conditions for the implementation of a new generation of computers. At each stage, this process implies setting up procedures which allow the analysis of the internal logic of the activities concerned. This progressively transforms the structures of firms and goes in the direction of an increased automatisation of decision procedures. Such an evolution tends to reduce the uncertainty of strategic decisions because it increases the control of the firm over the environment and in this sense 'internalises' it within the strategic scope of the firm (Simon, 1980). The introduction of more sophisticated means of storing, processing and communicating information tends to displace the limit which exists between the 'algorithmical' and the 'non-algorithmical' part of the firm. At each further step, as codification (algorithmisation) proceeds, discretionary decision capabilities can be applied to problems of higher complexity.[10]

Obviously this description of such a cumulative process is a simplification. In fact it may not be so continuous. Threshold effects appear each time that the algorithms underlying the functioning of a part (or the whole) of the firm can no longer react to the evolution of the environment. In a sense, organisational failures and bottlenecks in information accumulation are the equivalent to the physical limits in 'scaling up' in the standardised system of production.

(b) The control of the external environment: the emergence of the coordination function. Whether the potential future growth regime is regular or not, the cumulative character of the integration of information by successive 'algorithmisations' will not stop too soon. Flexible equipment is of no value without constantly renewed information. Each time the variety of products augments somewhere in the system, there will be a tendency to multiply the variety of the related components in order to respond to increasingly differentiated and specific needs. Consider again the case of advanced materials. In order to conceive and produce a growing variety of objects, the knowledge of *zones of compatibility* of different properties becomes of major importance. *Flexibility does not mean convexity of the properties space.* The ability to master potential combinations will be a future firm-specific asset among competitive firms in developed countries. Between equally flexible systems, the difference will be made by the possibility of stretching these zones of properties beyond the frontiers defined by standard data banks. It is in this respect that R & D activities will certainly become the most constraining (and rewarding) activity because it will be able to ensure an informational specificity, and thus be of comparative advantage to particular firms.

If this is so, it is necessary to ask what are the determinants of competitiveness in a complex system such as this one. Who will best control technical progress? How will the results be appropriated? Who will control the most strategic areas? The system constructor or the materials suppliers? The system constructor (car or aircraft industry, for instance) is well aware of the range in which constraints are compatible for a given technical object. The materials suppliers, traditionally associated with a number of user industries, dominate the complete spectrum of properties. Inter-industrial technological partnerships provide evidence that the question remains unanswered, and indeed that it probably does

not have any single answer. We are at the beginning of this transformation and thus it is too early to foretell the precise articulation of the new industrial system. One characteristic is, nevertheless, clear. The growing complexity of the system will make the coordination function very important both within and among firms.

In fact, there is a field in which the complex network has already been very important for some time, namely the space programme. Space technology does not have a unique scientific body of knowledge, as is the case with chemistry or electronics. Space technology is essentially organisational with no 'production' of its own. It is merely a kind of agency (e.g. the European Space Agency) which organises an international network of different industries. It is through a process of coordination and imposing of technical and organisational constraints (command delays, charge books, defined standards, etc.) that the space network weaves itself. The spillover of knowledge from the space departments to other technological and commercial fields comes from the nature of the participation in the immense technical and organisational coordination network.[11] According to their role in the network the firms do not develop the same learning profile, and thus the induced benefit categories are of a different qualitative nature. This example, we suggest, can be generalised. Hence, looking beyond the traditional dimensions of sectoral stratification and market structure, the role of the firm in the 'skill network' will become increasingly decisive for competitiveness.

Conclusion

The growing integration of information technologies, aided by the potentially tailor-made new materials, shifts the economic emphasis from capital-intensive to information-intensive production. We have tried to show how technological learning and the management of technological options differ in this context. If scale economies diminish in importance, inputs and outputs variety increase and transaction costs rise. To overcome transaction costs induced by the continuous redefinition of micro-markets and rising R & D costs, such 'Information-Intensive Production Systems' should prove their economic viability by finding some means of increasing efficiency. We have argued that, indeed, new technologies provide such a potential. Statically, the economies of scope of flexible equipment may compensate for information costs. Dynamically, a process of 'algorithmisation', that is, codification and routinisation of decision-making via more and more information-processing, may increase organisational efficiency, widen the strategic scope of the firm, and 'internalise' more control over the environment.

Notes

1. The advantage of the Al-Li alloys is that they do not require, up to a certain percentage of lithium, a costly redesign of both the aircraft and of the processing machines as in the case of composite materials. Therefore their diffusion does not imply high fixed costs. As we shall see, reconception thresholds are a very important feature in the competition among technological alternatives. Avoiding them, or at least postponing them, may have a large impact on the effectiveness of competing technologies. (See, on the issue of competing alternative technologies, the chapter by Arthur in this volume.)
2. 'Factors affecting the diffusion of technology', in N. Rosenberg (1972).
3. See Zuscovitch and Arrous (1984).
4. Problems raised by the interaction between user and producer are extensively discussed by Lundvall in Chapter 17.
5. Remember, however, that here we treat micro-market size and quasi-rent as given while they are equally determined by the nature of the income distribution. Income distribution, however, is partly determined in a macroeconomic setting, as are the optimality and welfare properties. An IIPS will certainly not have the same macro overall adjustment properties as the standard mass-production regime. Severe segmentation of markets, of labour and of goods would probably be less mean-dependent and more variance-dependent in all respects. Unions rightfully dread a structural deterioration in their power but also, more importantly, in the very capacity of workers to defend their basic rights once this segmentation is pushed too far. The only guarantee against exploitation is that the human capital is an essential component of the whole. In the same way that the product will be partly defined by the consumer so that he can expect to share consumer surplus, white- or blue-collar workers will also enter product definition. Apparently partnership is the name of the new game and everybody should be happy—except for those who are excluded, of course. The unemployed within and outside the developed countries will still call for standard solutions, provided that there will be solutions. Variety is the rich man's problem. Redistribution of resources will be even more needed than before in order to ensure minimal social integration.
6. For recent progress on selection processes in economics of technological change, see Nelson and Winter (1982) and Gibbons and Metcalfe (1986). In this book, this issue is discussed in the chapters by Silverberg, Metcalfe, Dosi–Orsenigo and Allen.
7. For a complete definition and analysis, see J. Panzar and R. Willig (1981).
8. See E. Bailey and A. Friedlaender (1982) for a formal definition.
9. This has also some implications for the organisation of the firm and particularly for its hierarchical structure. As Kay points out in Chapter 13, highly flexible structures, or what he calls 'organic systems', are better suited to meet the quickly changing environment of IIPS. 'Mechanistic systems', characterised by functional specialisation and formal hierarchical relationships, appear less efficient in managing dynamically fluctuating environments.
10. See Zuscovitch and Brendle (1985) for a detailed analysis of this process.
11. See Cohendet and Zuscovitch (1985).

References

Arrow, K. J. (1973), 'Information and economic behavior', in *Collected Papers of K. J. Arrow*, Vol. 4, *The Economics of Information*, Oxford, Basil Blackwell (1984).

Bailey, E. E. and Friedlaender, A. F. (1982), 'Market structure and multiproduct industries', *Journal of Economic Literature*, vol. XX, September, pp. 1024–48.

Cohendet, P. and Llerena, P. (1987), 'Flexibilité, complexité et intégration dans les processus de production', Actes du Seminaire International 'AMES', Paris, April, pp. 101–22.

Cohendet, P. and Zuscovitch, E. (1985), 'L'apprentissage techno-économique du programme spatial européen', SITEF, Toulouse, October.

David, P. A. (1975), *Technical Choice, Innovation and Economic Growth*, Cambridge, Cambridge University Press.

Gibbons, M. and Metcalfe, J. S. (1986), 'Technological variety and the process of competition', Conference on Innovation Diffusion, Venice, March 1986.

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

Panzar, J. C. and Willig, R. D. (1981), 'Economies of scope', *American Economic Review*, vol. 71, no. 2, May, pp. 268–72.

Rosenberg, N. (1972), 'Factors affecting the diffusion of technology', *Explorations in Economic History*, Fall.

—— (1982), *Inside the Black Box*, Cambridge, Cambridge University Press.

Simon, H. A. (1980), *Le nouveau management*, Paris, Economica.

Williamson, O. (1985), *The Economic Institutions of Capitalism*, New York, Free Press.

Zuscovitch, E. (1986), 'The economic dynamics of technologies development', *Research Policy*, vol. 15, pp. 175–86.

Zuscovitch, E. and Arrous, J. (1984), 'La diffusion intersectorielle des matériaux synthétiques', in P. Cohendet (ed.), *La Chimie en Europe*, Paris, Economica.

Zuscovitch, E. and Brendle, P. (1985), 'Informatisation: l'impact sur l'organisation des enterprises', *La Revue Française de gestion*, No. 51, mars–avril–mai.

12 Technological change and the nature of the firm*

David J. Teece

School of Business Administration, University of California, Berkeley

Introduction

Modern capitalist economies have a variety of organizational forms within which research and development is conducted. They include universities and government and private laboratories. Research and development laboratories differ in size, in the scope of scientific disciplines represented within them, and in the mechanism by which they are funded. These institutions both cooperate and compete to varying degrees and transfer know-how in and out.

The predominant mode of industrial research in the private sector, at least in the United States, is the integrated research organization, part of a business enterprise which engages in at least one other activity vertically related to research and development such as manufacturing, marketing, distribution, sales and service. This chapter focuses on this particular component of the research infrastructure of modern American capitalism. It attempts to explain the reluctance on the part of innovating enterprises to rely on external research facilities to procure new products and processes via the market.

In view of the historical reluctance of firms to contract for technology, the sudden and recent rise in external acquisition activities by certain US corporations warrants an explanation. Relatedly, the 'hollowing' of the corporation—that is, the outsourcing of components and in some case whole systems—is explored with respect to possible ramifications for the appropriability of returns from innovation.

A second and subsidiary theme explored in this chapter is the relationship between technology and technological change, and the growth or diversification activities of the business enterprise. Given the current state of economic theory, one should be equally surprised by the diversification as by the coherence—that is, the tendency for firms *not* to be pure conglomerates with their activity randomly spread across a variety of product lines—of the modern corporation. It is hypothesized here that a good deal of the coherence of the corporation can be understood in terms of techno-

*I am especially grateful to Richard Nelson, Giovanni Dosi, Sidney Winter, Gary Pisano and Oliver Williamson for helpful discussions that have shaped my thinking on the issues addressed in this chapter.

logy, technological change and the differences between technologies in their managerial requirements.

Historical perspective on the organization of private-for-profit R & D activities

During the late nineteenth century and the first half of the twentieth century, American manufacturing firms bought an increasing share of R & D in-house. Previously, practically all of it had been conducted outside of the firm in stand-alone research organizations. Thomas Edison's industrial research laboratory in Menlo Park, New Jersey, was one such structure, and from it flowed the light bulb and many other inventions.[1] Even as late as 1945, there were hundreds of such organizations employing over 5,000 scientists and engineers.

Throughout the early decades of the twentieth century, however, the independent, stand-alone labs were in relative decline (see Figure 12.1); and during certain decades, they probably actually declined in absolute number. In 1911, for instance, Arthur D. Little organized for General Motors a laboratory for materials analysis and testing. But the main component of G.M.'s research organization came from an independent lab—the Dayton Engineering Laboratories Company—which was absorbed by G.M. after being organized by Charles Kettering and E. A. Deeds (Sloan, 1964).

Table 12.1: Employment of scientific professionals in independent research organizations as a fraction of employment of scientific professionals in all in-house and independent research laboratories, 1921–46

1921	15.2%
1927	12.9%
1933	10.9%
1940	8.7%
1946	6.9%

Source: Mowery (1983, Chapter 2).

This is not to imply that contract research and in-house research are substitutes. Mowery (1983) has suggested that they were complements, in the sense that as in-house research facilities grew in size and number during 1900–40 they also developed as the primary clients for the stand-alone research organizations. This may indicate that they were subcontractors bearing a vertical relationship to the in-house labs. Firms without in-house laboratories, moreover, used contract research only for the simplest types of research projects (Mowery, 1983, p. 363), a characteristic still evident today (Teece and Armour, 1977, p. 56).

Mowery's case studies of Arthur D. Little (founded 1896), the Mellon Institute (1911), and Batelle (1929) are instructive. Mellon's contract research was primarily concerned with the improvement of existing processes or the utilization of by-products. Nearly 25 per cent of Batelle's projects undertaken during the period 1929–40 were analyses or tests of metals, minerals or coal (few mining firms had in-house labs). Chemical analyses were a mainstay of ADL's activities. The evidence seems to indicate that the independent research organizations did not engage significantly in new-product development and did not offer a wide menu of contract research services.

In-house research thus came to be the dominant mode for supporting corporate research in America, for small as well as for large organizations. By the 1970s, there were very few stand-alone research organizations, and these typically performed a very limited kind of research. In the petroleum industry, only one such firm—Universal Oil Products—remained by 1970, and it has subsequently lost its stand-alone status. Teece and Armour (1977, pp. 56–7) noted the rather narrow range of research activities that were conducted under contract—typically, those where the research objectives are simply and obvious, and where the risks are low.[2]

The integration of R & D with production

Contractual analysis

The internalization of research and development warrants theoretical explanation. Why is it that in the modern capitalist corporation R & D generally nestles in close to marketing and manufacturing? Put differently, why do non-market modes rather than market (contractual) modes appear to dominate as a mechanism for securing the output of research establishments? After all, Stigler (1956, p. 281) has remarked that, 'We may expect the rapid expansion of the specialized research laboratory which sells its services generally. The specialized laboratories need not be in the least inferior to captive laboratories.' In order to explore these matters, a stylized organizational framework is assumed for an industry experiencing technological change (Figure 12.2). The figure shows the kinds of transactions/interactions which must exist between the organization if new technology is to be developed and implemented.

If one begins with the premise that there are gains associated with organizational specialization and that markets provide workable mechanisms for linking organizations, then Stigler's presumption that the stand-alone lab supported by contracts would outperform in-house labs naturally follows. Indeed, in most advanced industrial economies one observes a considerable amount of defense-related research being procured via contractual mechanisms. At the same time, most firms in industries experiencing rapid technological change have in-house R & D capabilities. An exploration of the relative efficiency properties of the two modes thus appears to be warranted.

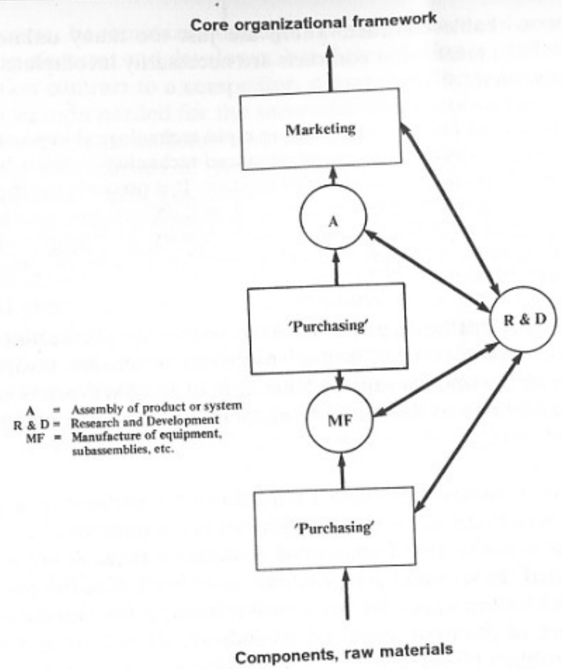

Figure 12.1

Two sets of factors are relevant. The first set of factors relates to economies of scale and specialization. The second set of factors relates to the workability of contractual mechanisms. Both issues will be examined from the perspective of the relative merits of A (assembly) or MF (manufacturing) being integrated into R & D. The contractual issues are essentially the same whether it is the assembler or the manufacturer purchasing R & D services.

If a purchaser (such as A or MF) of R & D services is comparing in-house and contract research alternatives, and if the R & D activity in question involves significant scale economies or capabilities which the purchaser does not possess, then standard microeconomic analysis would indicate that contracting for R & D services from an established low-cost provider will be the superior alternative, just as it would be for any other service or component. However, a more sensitive analysis reveals that contractual mechanisms for procuring R & D services are shot through with hazards, so that the costs and practicality of relying on the market for R & D services will be tightly circumscribed in many important circumstances. This is because contracts encounter difficulties as the degree of uncertainty increases. The greater the uncertainty, the more difficult it is to specify a workable contract. There are just too many unknown contingencies, which means that contracts are necessarily incomplete. According to one observer:

> it is inherent in an industry experiencing rapid technological improvement that a new product, incorporating the most advanced technology, cannot be contracted for by detailed specification of the final product. It is precisely the impossibility of specifying final product characteristics in a well-defined way in advance which renders competitive bidding impossible in the industry. To attempt such a specification would, in itself, constitute a serious impediment to technical progress. [Miller and Sawyers, 1970]

Contractual difficulties are not, however, limited to specification problems. There are disclosure and 'lock-in' problems which are invariant to the mode of contracting chosen, whether it is of the fixed-prices or the cost-plus variety. Each of these two fundamental modes of contracting is now explored further.

Fixed-price contracts. The use of fixed-price contracts to procure new products which are state-of-the-art or beyond exemplifies the difficulties associated with the use of unassisted markets. A number of problems can be identified. First, there are problems associated with the production of precontract information. In competitive bidding for complex contracts, conveyance of information at the precontract stage is likely to be a substantial problem (Goldberg, 1977). Formulating specifications will require interaction with the potential providers at the precontract stage which will be both time-consuming and costly. The costs of transferring the information will influence both the relative efficacy of alternative bidding mechanisms and the nature of the output itself. They will also, of course, influence the relative merits of competitive bidding versus vertical integration.

A second problem the purchaser faces in relying on markets is obtaining the appropriate level of protection of proprietary information—his own and that of the potential suppliers. Conveying accurate information to potential bidders can decrease the likelihood that some valuable trade secrets will be protected. Likewise, a solicitation which requires suppliers to reveal confidential information might induce those suppliers to forgo the bidding or demand costly safeguards.

A third problem is 'lock-in'. That is, there are limited options for changing developers, as well as limited options for using a supplier (other than the original developer) to perform subsequent production. This is due largely to the fact that in order to secure competitive bidding at the production stage, detailed manufacturing drawings and specifications must be created. Yet drawings executed for development purposes are often necessarily incomplete. Short-cuts may be taken and engineers often rely upon verbal communication with production foremen to ensure that the proper manufacturing sequences and tolerances are satisfied. The tacit component (Teece, 1981; Nelson and Winter, 1982) is often high. Errors

might, of course, be minimized through intimate contact and cooperation between manufacturer and developer, but if the developer has lost out on the production contract to a competitor, it may be difficult to achieve the kind of cooperation needed for the successful transfer of technology from the developer to the manufacturer/producer.

Of course, the transfer of production away from the developer to another enterprise does not remove the 'lock-in' problem—it merely transfers the dependence from one potential supplier to another. This can be relieved, at least in part, by second-sourcing strategies, but this is also expensive as manufacturing drawings and know-how must be conveyed by the original producer to the new second-source producer with attendant costs and delays. Furthermore, if economies of scale are present, considerable production cost savings may be sacrificed by a dual-sourcing strategy. Similarly, when learning is an important factor, experience curve advantages may be lost by second sourcing. Also, quality control and standardization are more difficult to achieve when multiple supply sources are involved.

The above problems are softened once the design of a new product has been stabilized. It may then be possible to rely on contractual mechanisms to achieve efficient supply. When specifications define in detail the product to be procured, the buyer has a better chance of assuming that the contractor delivers what was promised, and the contractor can in turn ensure that it will be asked to deliver no more than was promised.

Cost-plus contracts. Because of the above problems, and because R & D costs are subject to enormous uncertainty, it is hazardous if not impossible to determine the price for a product which has yet to be created. Unless both parties are risk-neutral, there may be a reluctance to enter fixed-price contracts. An alternative is the cost-plus contract.

However, the cost-plus contract is unlikely to be superior to integration, except where a one-shot transaction is contemplated—in which case it may not be cost effective to build the requisite internal capabilities. The reason is that the cost-plus contract has only weak incentives for least cost performance. In fact, in order to ensure that the gross abuses of this mechanism do not occur, it may be necessary to set up an administrative structure which replicates many of the features of vertical integration.[3]

For instance, in 1955 the Consolidated Edison Company of New York used a cost-plus contract with Babcock and Wilcox for the development and construction of a 235,000 kW atomic energy plant where the relationship involved major technological uncertainties. But the relationship was hardly arms-length. Consolidated Edison exerted detailed supervision over Babcock and Wilcox, auditing all their invoices, approving all engineering changes, and authorizing any variations that might affect the level of cost. This was a highly detailed control by the buyer of the contractor's activities.[4] In short, in order to monitor cost-plus contracts, an administrative mechanism must be set up which is tantamount to vertical integration.

Indeed, when the procurement of technologically advanced products and systems is contemplated, the key organizations responsible for delivering advanced products and systems will need to have an in-house research and development capability.

Contracting for R & D exposes the enterprise not only to an R & D 'lock-in' in that valuable knowledge may be required which will give the R & D organization an advantage with respect to subsequent R & D contracts, but to a manufacturing 'lock-in' as well if the R & D service procured involves the design of equipment or components which will then be embedded in a product (or service) which the procurer is manufacturing. The lock-in problems stem, at least in part, from the costs of technology transfer from the developer to the producer. The developer thereby obtains a first-mover advantage in production which can be used to advantage by the developer if the developer is also a potential supplier at the production stage. This is true for both fixed-price and cost-plus contracts. The reason is that during development activities a considerable amount of tacit knowledge is acquired in a learning-by-doing fashion. If the development work is extensive and costly to replicate, then the procurer is exposed to a 'lock-in' in the sense that subsequent production cannot be contracted in a fully competitive fashion. The developer will have acquired a first-mover advantage and may be able to price subsequent production above long-run costs because of the advantage it has acquired, relative to its rivals, at the development stage. The phenomenon has been commented upon in the context of weapons acquisition:

> Consider two firms, A and B. Firm A is the sole developer of the weapon; both firms are capable of producing it. Now, since firm A has developed the weapon, it is reasonable to assume that A has garnered some knowledge from the development process which will be useful in the production process. Consequently, the original developer, in this case A, is capable of producing the first x units at a lower cost than any other producer. [Arditti, 1968, p. 320]

Arditti presents supportive data from the airframe industry. In a different context, data have also been assembled which suggests that the switching costs are especially high before the new preoduct has gone into production. The available evidence strongly supports the contention that it is knowledge and the high cost of its transfer which yields the advantage to the first mover. Furthermore, the less codified is the relevant know-how, and the closer it is to the state-of-the-art, the more costly it is to transfer (Teece, 1977). In short, when the amount of development activity involved is large, the procurer may well provide the supplier with a non-trivial first-mover advantage if the work is performed at the procurer's expense. In these circumstances, the procurer can avoid the 'lock-in' problem via the vertical integration of production.

Integration between research and the user. Contractual analysis makes it clear that integration between the manufacturer and R & D is usually

necessary both because of the difficulties associated with specifying the R & D services which are to be procured as well as subsequent lock-in. In the above discussion, it was assumed that the organizational unit desiring R & D services had a clear perception of what was needed, even if that perception could not be translated into workable specifications.

The identification of market requirements is, however, a complex process itself. An essential feature of successful innovation is that it must be responsive to user needs. The available evidence indicates that successful attempts at innovation are distinguished frequently from failures by greater attention to the understanding of user needs. Innovation involves a complex series of events. A number of interfaces must be crossed in the process of technological innovation. Each interface becomes a potential barrier to innovation unless spanning mechanisms are put into place. At least three different kinds of spanning or gatekeeping functions are commonly acknowledged as critical to innovative success (Roberts, 1979): the technical 'gatekeeper', the market 'gatekeeper' and a manufacturing 'gatekeeper'. The technical gatekeeping function bridges the organization to the scientific community at large. The market gatekeeper function must understand what competitors are doing, what the regulators are up to, and what is happening with respect to changes in the customer marketplace. The person or persons communicating these sources of information to the R & D environment is a critical contributor who keeps the technical organization on target towards the kinds of activities that will eventually be successful in the marketplace (Roberts, 1979, p. 27). The manufacturing gatekeeper function develops understanding of the real and hard-nosed environment of the manufacturing plant so as to keep R & D up to date on the realities of materials, of assembly processes, and to keep marketing and R & D up to date on the cost of doing things different ways. This function helps ensure that what gets designed and developed in R & D is targeted towards producibility at a cost sufficiently low to generate meaningful volume and profitability.

These functions are best performed by specialists who understand each other's problems and needs, who share common objectives, and who can collaborate and exchange the information that each needs freely and without corporate or proprietary barriers. Integration clearly facilitates the activities of such specialists. Intra-organizational boundaries are typically more permeable than market boundaries—in part because secrecy is not jeopardized, and a common internal language can be employed. The existence of a common coding system and the attendant dialogue among organizations facilitates both technology transfer and the formulation of appropriate research objectives. As a result, the research activity is likely to be better directed and hence more productive.

Cumulative learning and spillovers in R & D

The analysis presented in the previous section sheds light on considerations which very often compel integration of the R & D activity or grounds of contractual efficiency. In most cases, however, there are additional factors which are not fully elucidated by focusing on contracting and technology transfer issues alone, although in a theoretical sense they can possibly be thought of in contracting terms.

As was indicated above, the phenomenon of 'lock-in' is likely to characterize the relationship between the R & D provider and the unit utilizing the results of the R & D process. 'Lock-in' has its roots in high switching costs, often due to the fact that much of the technology generated by R & D activities is of the tacit kind, and this is costly to transfer.

This tacit knowledge, moreover, tends to be cumulative, which is all the more reason why it is often desirable to deepen and stabilize the relationship among the buyer/user and the seller/provider. As Nelson and Winter (1982) have explained:

> In many technological histories the new is not just better than the old; in some sense the new evolves out of the old. One explanation for this is that the output of today's searches is not merely a new technology, but also enhances knowledge and forms the basis of new building blocks to be used tomorrow. [pp. 255–6]

Because the knowledge acquired in the course of one project often has implications for the next round of R & D projects, it is important that the entity which is sponsoring the R & D activity keep a close liaison with the R & D unit, not only to access valuable technology and possibly firm-specific knowledge that the R & D unit will have generated, but also for reasons of preventing this from 'spilling over' to competitors. Spillover would almost certainly occur if the R & D unit was free standing.

This 'neighborhood' characteristic of discovery not only explains why firms need to keep the R & D activity in-house; it also explains why there is a certain natural trajectory associated with a firm's *de novo* product migration and diversification activity. This is the topic of the next section.

Determining product boundaries with technological change

The firm's 'core business'

A firm's core business, it can be argued, stems from the underlying natural trajectory embedded in the firm's knowledge base. New product development thus usually proceeds 'close in' to previous successes. A wave of improvements, often dramatic in their significance, may follow the introduction of a major new technology. The desirability of hunting for improvement depends, of course, upon the commercial promise which first commercialization may have yielded or at least signalled. Previous commercialization histories may indicate the most promising technological neighborhoods to explore in terms of market acceptance. Sometimes, however, there may be a kind of inevitability to the direction of search, driven by what Rosenberg (1969) has referred to as 'technological imperatives'.

A change in technological regime—where regime is defined by the convergence of engineer beliefs about what is feasible or at least worth attempting—or the simultaneous coexistence of several related technological regimes (as with both CMOS and NMOS technologies in semiconductors) may soften technological imperatives, making them less path-dependent. Of course, technological discontinuities may blow path dependencies asunder.

There are important implications for economic theory, for the organization of research, and for the organization of economic activity more generally. First, because promising areas of research inquiry lie 'close in', and because a set of production/manufacturing activities are typically implied by a particular research focus, a firm's 'core business' (or possibly core businesses)—by which is meant the set of competences which define its distinctive advantage—can be expected to display a certain stability and coherence. Path dependencies inherent in technological progress can be expected, at least partially, to drive the definition of a firm's capabilities, and therefore the businesses in which it has a comparative advantage.

Buttressing path dependency as a limiting factor with respect to a firm competency are the set of organizational routines which develop once a particular research endeavor bears fruit. Whereas the creative part of research is at least partly *ad hoc*, routines characterize efficient post-development behavior in production, marketing, distribution and sales. A routine is defined by putting a skill, or set of skills, to use in a particular or distinctive environment in a repetitive way. As mentioned earlier, path dependencies define the neighborhoods/environments in which skills can be most productively applied. Hence, a firm's initial point of entry in a technological regime, and the trajectories/paths which are initially selected, will define in large measure the kinds of competences that the firm will generate, and the products it will develop and commercialize. After first commercialization, a set of routines will develop which will lead to a deepening of competencies in certain areas. The skilful performance of organizational routines provides the underpinnings for what is commonly thought of as the distinctive competence of an enterprise. 'These people at company *X* are good at *Y*' summarizes views which outsiders may develop with respect to these competences. These competences, coupled with a modicum of strategic vision, will in turn help define a firm's core business. A firm's core business is necessarily bounded by particularized competences in production, marketing and R & D. Employees will tend to form natural teams, where regrouping is difficult and the ability to absorb new members limited.

Routines, since they cannot be codified, must be constantly practised to exhibit high performance. This in turn implies that the firm must remain in certain activities in which short-run considerations would indicate that abandonment is desirable. Put differently, core business skills need to be constantly exercised to retain corporate fitness.

Because a firm's learning domain is defined in part by where it has been, and the technological imperatives and opportunities which that implies, it is readily apparent that a firm has a limited but by no means a non-existent ability to change its business. The products it can produce and the technologies it employs are highly path-dependent, at least at the level of an individual business unit. At the level of the corporation, more can be done, but this typically involves entering the corporate control market (i.e. buying and selling businesses) and not the market for factors of production.

Implications for the theory of the firm are apparent. Except by entering the market for corporate control, profit-seeking firms have limited abilities to change products and technologies. The notion of a smooth, twice differentiable production function would appear to be at odds with the conceptualization of the firm outlined above. In addition, economics of scope would appear to be constrained by the limited ability to apply routines across different product and technological environments/neighborhoods. 'Related' diversification would appear to be feasible so long as it is consistent with the underlying path dependencies and/or imperatives, a matter which is to be discussed in more detail later.

The analysis has so far assumed relative stability with respect to technological regimes. Suppose, however, that the firm experiences a technological discontinuity which obsoletes its skills, and possibly renders its downstream assets valueless. In these circumstances, established firms will be lacking in many of the relevant research competences. However, 'downstream' competences, particularly in sales and distribution, more often than not are still relevant to the new technological regime.

In these circumstances—that is where the technology necessary for survival lies distant from the neighborhood of the firm's traditional research inquiry—it may be extremely difficult to utilize existing in-house research competences within the new paradigm. This is because of the path dependencies noted earlier. Accordingly, the relevant competences may have to be purchased *en masse*, or technology transfer programs must be employed to educate existing personnel in the assumptions and logic of a new paradigm. In these circumstances, in-licensing and collaboration with the organizations (typically universities or new business firms) responsible for pioneering the new paradigm will be common.

Incumbent firms can thus be expected to display permeable boundaries when technological regimes shift, unless of course incumbents have been responsible for the shifts. However, if the know-how in question is not protected by intellectual property law, then the collaboration at issue is likely to be more in the form of imitation rather than in licensing. However, if the technology has a large tacit component, know-how licensing may still be necessary. Needless to say, transactions-cost issues also enter the equation, with collaboration more likely less difficult than contractual problems.

Multiproduct diversification

Technological change is often driven, so it seems, by certain imperatives in a trajectory which, considered in light of the firm's market-entry strategy, helps define the firm's 'core business'. However, the diversity of application areas for a given technology are often quite large, and the possibility of applying the firm's capabilities to different market opportunities is often available, especially after growth opportunities in existing markets are exhausted.

Suppose application areas outside of the core business do in fact open up. The question arises as to whether potential scope economies deriving from the application of generic know-how in new markets add more to the innovating firm's value if they are served through licensing and related contractual arrangements to unaffiliated firms who then serve the new product markets in question, or by direct investment, either *de novo* or by merger/acquisition. This is an important question, the answer to which ought to help shape a positive theory of the scope of the firm's activities.

Whether the firm integrates or not is likely to depend critically on four sets of factors:

1. whether the technology can be transferred to an unaffiliated entity at higher or lower cost than it can be transferred to an affiliated entity;
2. the degree of intellectual property protection afforded to the technology in question by the relevant statutes and laws;
3. whether a contract can be crafted which will regulate the sale of technology with greater or less efficiency and effectiveness than department-to-department or division-to-division sales can be regulated by internal administrative procedures;
4. whether the set of complementary competences possessed by the potential licensee can be assessed by the licensor at a cost lower than alternatives. If they are lower, the available returns from the market will be higher, and the opportunity for a satisfactory royalty or profit-sharing arrangement accordingly greater.

These matters are explored in more detail elsewhere (Teece, 1980, 1983, 1986). Suffice to say that contractual mechanisms are often less satisfactory than the alternative. Proprietary considerations are more often than not served by integration, and technology transfer is difficult both to unaffiliated and affiliated partners, with the consequences that integration (or multiproduct diversification) is the more attractive alternative, except where incumbents are already competitively established in downstream activities, and are in a position to render *de novo* entry by the technology-based firms unattractive. Hence, multiproduct firms can be expected to appear as efficient responses to contractual, proprietary and technology transfer problems in an important set of circumstances. Mixed modes, such as joint ventures and complex forms of profit-sharing collaboration, will also be common according to how the set of transactions in question stacks up against the criteria identified above.

Vertical integration

Technological change also has implications for the vertical structure of the business enterprise, and the level of vertical integration also has implications for the rate and direction of technological change expected to characterize the business enterprise. However, the literature linking the rate and direction of technological change and the boundaries of the firm is still in its infancy.

Economic historians have long suggested that there may be links. For instance, Frankel (1955) has argued that the slow rate of diffusion of innovations in the British textile and iron and steel industries around the turn of the century was due to the absence of vertically integrated firms. Kindleberger (1964) has gone so far as to suggest that the reason why West Germany and Japan have overtaken Britain may be due to 'the organization of [British] industry into Separate Firms dealing with each other at arm's length'. This 'may have impeded technological change because of the possibility that part of the benefits of that change would have been external to the separate firms' (pp. 146–7). General Motors' early dominance in the diesel electric locomotive industry has also been attributed to the fact that it was integrated into electrical supply while its competitors were not (Marx, 1973). Clearly a systematic exploration of the relationship between technological innovation and enterprise boundaries is needed. This can be done by comparing the properties of integrated structures with non-integrated structures which rely on arms-length contractual relations to achieve the requisite degree of coordination.

For present purposes, it is useful to distinguish between two types of innovation: autonomous (or 'stand-alone') and systemic. An autonomous innovation is one which can be introduced without modifying other components or items of equipment. The component or device in that sense 'stands alone'. A systemic innovation, on the other hand, requires significant readjustment to other parts of the system. The major distinction relates to the amount of design coordination which development and commercialization are likely to require. An example of a systemic innovation would be electronic funds transfer, instant photography (it required redesign of the camera and the film), front-wheel drive, and the jet airliner (it required new stress-resistant airframes). An autonomous innovation does not require modification to other parts of a system for first commercialization, although modification may be necessary to capture all of the advantages of the innovation in question. The transistor, for example, originally replaced vacuum tubes and the early transistor radios were not much different from the old ones, although they were more reliable and used much less power. But one did not have to change radio transmission in order to commercialize transistor radios. Another example would be power steering—the automobile did not have to be redesigned to facilitate the introduction of this innovation, although it did permit designs which placed more weight over the front wheels. A faster microprocessor or a larger memory would be further examples.

When technological interdependencies are important, it is likely that the commercialization of an innovation will require new investments in several different parts of the industry or the system. Thus, suppose that a cost-saving (equipment) innovation has been generated which can enhance efficiency if successfully introduced into an industry, and suppose that introduction into one part requires that complementary investments be made in other parts. If the subparts are independently owned, cooperation will have to be obtained in order for the innovation to be commercialized.

There are two powerful reasons why common ownership of the parts will speed both the adoption and the subsequent diffusion of the innovation. Where there are significant interdependencies, introduction of an innovation will often result in differing benefits and costs to various parties. This effect makes it difficult if not impossible to coordinate the introduction of such an innovation. While a system of frictionless markets could overcome this problem—the firms obtaining the benefits could compensate those incurring the costs so that the introduction of the innovation would not depend on the degree of integration in the industry—it is commonly recognized that it may be extremely difficult to engineer a workable compensation agreement, in part because all relevant contingencies are not known when the contract would need to be drawn up.

Therefore, in the absence of integration, commercialization can be slowed or completely stalled. Considerable cost disparities can open up between old and new methods, yet the new method may not be implemented because the individual parties cannot agree upon the terms under which it will be introduced. There can be a reluctance on the part of both parties to make the necessary investments in specialized assets, and to exchange information about each other's needs and opportunities—even if cooperation would yield mutual gains, and certainly if the gains will go to one party at the expense of the other. Hence, in the absence of integration, there can be a reluctance on the part of one or more of the parties in an industry to develop or commercialize a systemic innovation requiring the participation of two or more firms.

To summarize, it is hypothesized that integration facilitates systemic innovations by facilitating information flows, and the coordination of investment plans. It also removes institutional barriers to innovation where the innovation in question requires allocating costs and benefits, or placing specialized investments into several parts of an industry. In the absence of integration, there will be a reluctance on the part of both parties to make the necessary investments in specialized assets, even if this would yield mutual gains. One reason is that both parties know that the exercise of opportunism might yield even greater benefits to one of the parties. Hence, in the absence of common ownership of the parts, there will be reluctance on the part of one or more of the parties to adopt a systemic innovation.

Comprehensive evidence with respect to the above concepts has yet to be assembled. The only statistical test performed to date relates to the petroleum industry (Armour and Teece, 1978). These findings indicated that firm and R & D expenditures for basic and applied research in the US petroleum industry, 1951–75, were statistically related to the level of vertical integration which the enterprise possessed.[5] Anecdotal historical evidence is surveyed below.

According to Frankel (1955, pp. 312–13), the lack of vertical integration in the British iron and steel industry hindered the introduction of technical innovation in the latter part of the century because the innovations in question displayed interrelatedness. In the 1860s, 1870s and 1880s, a number of technical changes evolved which offered the prospect of substantial economies in various phases of production. New and commercially feasible methods of steel ingot production were possible through introduction of the Bessemer converter and the open hearth and electric furnaces; advent of the rolling mill made possible phenomenal savings in the shaping and finishing of steel products; the optimum size of the blast furnace was greatly increased; and significant fuel economies all along the line became possible. Most important, no single one of these changes would yield its potential savings in full except in conjunction with the others, thereby compelling coordinated design of investment and the assembling of the various parts of the industry at a single location. The economies attainable from coordinated design and centralization derived from several sources: from a reduction in transport and handling charges; from higher capacity utilization; from easier and more accurate control over product quality; and from fuel economies. But to attain this, these had to be in such proportions and on such a scale that the whole plant could work effectively and economically. There had to be a sufficient number of coking ovens, blast furnaces and steel furnaces both to keep one another employed and to meet the requirements of the rolling mills. The technical changes that made possible the economies in question materialized in an era when the ownership structure in Britain had crystallized in a pattern inconsistent with the new technological need for integrated operation. As a result, in part at least, those changes had hardly begun to be assimilated in the United Kingdom by the end of the century.

Frankel (1955, pp. 313–14) provides further supporting evidence for the general proposition from the textile industry. He argues that the failure of the British to put the automatic loom into place in the cotton industry was due to the lack of vertical integration. The loom was first introduced around the turn of the century, and by 1914 represented 31 per cent of the looms in operation in the United States. By 1919, the figure was 51 per cent and by 1939, 95 per cent. In Britain, by 1939, only 5 per cent of the looms were automatic. Introduction of automatic looms demanded more than replacement of one machine by another. For a great many firms it required redesign of the weaving shed—often its complete rebuilding, strengthening of flooring, elimination of pillars, and respacing of machinery. It called for equipment and method changes in the preliminary processing of yarn; idiosyncratic investments were required in both the spinning and weaving. Furthermore, the innovation created a need for product simplification and for a change in the traditional practices of converters in allocating orders to weavers, so that long runs of given fabric types could be attained. Certain of the interconnections were external to the firm. Frankel (ibid.) concludes that 'Their presence, together with those internal to the firm, constitute a part of the explanation for Lancashire's continued reliance on the old power loom', while the automatic loom diffused very rapidly in the vertically integrated US industry.

Other historians share this perspective. Kindleberger has studied the reasons for the failure of the British railroads to abandon the 10-ton coal wagon in favor of the more efficient 20-ton wagon. One hypothesis considered is that the retention of the 10-ton car was due to the impossibility of changing it without also modifying terminal and switching facilities, which might have made the cost prohibitive. Another hypothesis—and the one which Kindleberger comes to favor—is that the lack of integration blocked the adoption of the larger cars. (The British railroads were peculiar in that the coal wagons were owned by the coal mines and not the railroad companies.) In this regard, it is of interest to note that in this period two of the British railroads—the Great Western and the Great Eastern—adopted 20-ton wagons for their own use for locomotive coal as early as 1897; the Northeastern had used 20-ton, bottom-discharge mineral wagons for iron ore and 40-ton bogie wagons since the beginning of the century. But as Kindleberger (1964) notes:

> These wagons were all owned by the railroads, not the coal or iron companies. The Great Western Railway failed, however, to persuade colliery owners to change to a larger wagon when it offered a rebate of 5 percent on freight cars for coal in fully loaded 20-ton wagons in 1923 and, in 1925, reduced charges on tipping and weighing these wagons. Only 100 came into use. [p. 143]

Kindlberger (1964) concludes that the reason for the slow rate of diffusion was institutional and not technical. In short, it stemmed from the absence of vertical integration.

> Technical aspects of interrelatedness do not seem to have held up the movement to more efficient size, either through making such a change uneconomic because of the enormity of the investment required or by adding amounts too great for any one firm to borrow. The sums involved were not large, and railway finance was rarely a limiting factor in the period up to 1914. Private ownership of the coal cars by the collieries, on the other hand, posed a type of interrelatedness that was institutional rather than technical. [ibid.]

Kindleberger (1964) does not advance a satisfactory explanation of why market mechanisms could not achieve the requisite coordination, but he hints at the difficulties of devising a mechanism for sharing the gains from the innovation.

> For the railroads to guarantee new low rates on larger wagons would have been to take all the risk of the new investment, and that of the collieries, on themselves. To indicate that they would consider new and lower rates only if these were justified by

operating economies would assign the risk to the collieries. An attempt to apportion the risk between the two would have been equitable but not likely to arouse much enthusiasm. [ibid.]

An alternative reason might well be that the collieries had no guarantee that the lower rates would prevail. In the absence of competition from other railroads or modes offering competitive rates, the railroads could raise the rates again once the collieries had made the requisite investments. Furthermore, to the extent that the economies could not be captured until substantially all of the mines had adopted the larger cars, the rate reduction offered may have been insufficient to entice the mines to make the needed investments.

A more general impediment to innovation has been identified in the British distribution system. British industry in the nineteenth century displayed very little in the way of forward integration—there was a layer of merchants between the manufacturer and the final customer. While enabling specialization economies to be obtained in a static market, Kindleberger suggests that 'the separation of selling from production may have the drawback of slowing down technical change by imposing barriers of communication between the ultimate customer and the producer' (1964, p. 148). Furthermore, 'it may be significant that the woolen industry, which did much better than cotton in maintaining its rate of technical change, moved to direct trading' (ibid.). His final conclusion is that:

There is the distinct possibility, whose complete demonstration would require a separate book, that the merchant system bears a significant share of the responsibility for slowing down technical change because it renders a large proportion of the benefits of technical change external to the firms that must effect or sell it. [p. 149]

A more contemporary episode which is instructive is how General Motors' integration between locomotive manufacture and electrical equipment supply facilitated GM's commercialization and subsequent success with its diesel electric locomotive program. The diesel locomotive revolution began in 1934 when General Motors produced the first diesel electric passenger locomotive. The domestic diesel electric locomotive building industry currently consists of GM and General Electric. GE did not enter the industry as a fully integrated builder, producing both the diesel engine and the electrical components, until the early 1960s. Previously, GE had confined itself to the electrical equipment supplier end of the business. The electrical equipment in a diesel electric locomotive represents approximately one-third of the total locomotive cost. GM, on the other hand, originally purchased its electrical equipment from others but, in the late 1930s, it integrated upstream into electrical supply. From the 1930s until the 1960s when GE entered, GM was the only integrated producer. The producers of steam locomotives, including Alco, Baldwin, Lima-Hamilton and Fairbanks Morse, had all abandoned locomotive production by the 1960s.

One effect of GM's integration was the elimination of duplicate personnel. The locomotive builders maintained in-house electrical engineering staffs even though they sourced their electrical equipment externally. These in-house staffs were apparently maintained because of the problems of information exchange associated with market contracts.

The maintenance of in-house electrical engineering staffs facilitated the reception of technical engineering and price data, and reduced the extent of undetected data distortion. These market transactions costs arise from the small number of producers and the uncertainty that marked the locomotive building industry . . . Uncertainty and market transactions costs are greater when technology is changing rapidly. In this case, advance specification of final product design and cost was impossible. [Marx, 1976, pp. 45–6]

Another advantage of integration was that it eliminated costly disputes with respect to warranty responsibility. Haggling and the opportunistic interpretation of contractual ambiguities marked the Alco–GE relationship, which was often described as 'antagonistic' by railroad executives. A particularly difficult problem was the identification of the source and responsibility for engine failure. Because of the systems nature of the technology, it was never very clear whether the problem was mechanical or electrical. Locomotives requiring service were frequently shuffled back and forth between the building and electrical equipment supplier. Because of the costliness of locomotive downtime, the availability of timely after-sales service is a critical factor in procurement decisions. The railroads frequently cited the superiority of General Motors' post-sales service as an important factor in its sales record.

Thus, by facilitating coordination and timely post-sales service, vertical integration appears to have contributed to the market success of the GM diesel electric locomotives. Furthermore, integration also facilitated rapid technological development. According to Marx (1976), one of the biggest problems with the electrical equiment suppliers was:

their commercially cautious and slow rate of development. The electrical suppliers were more risk averse than the builders because of different commitments and alternatives. This produced different rates of development for mechanical and electrical equipment and, because of the interdependence, technological bottlenecks. [p. 47]

Vertical integration also helped solve appropriability problems. The locomotive builders which were not vertically integrated typically funded a portion of the electrical equipment supplier development cost, depending on the exclusivity of the work. The identification, sharing and pricing of intellectual property generated by such development activities turned out to be difficult and costly.

Contract terms notwithstanding, the indirect benefits of contract execution (accumulation of knowledge, staff development, and the like) accrue to the supplier, and enforcement through secrecy is seriously impaired. General Motors also needed to exclude rivals from sharing in the results of its own in-house mechanical and electrical research. The exclusion problem was difficult because of the close technical contact required between builders and electrical equipment suppliers, who also manufactured for competing builders (Marx, 1976, pp. 49–50). Vertical integration, by harmonizing the divergent interests of the locomotive and electrical equipment producers, helped overcome many of the problems associated with the exclusivity and appropriability of the technology. Attention could then be focused on getting the job done at the lowest cost.

In conclusion, it appears that GM's integration into electrical equipment supply reduced costs by internalizing market exchange under circumstances (uncertainty, technological interdependence) which generated significant contractual difficulties. This integration also stimulated the pace of product development by promoting harmonious information exchange. The experience with vertical integration in the diesel electric locomotive building industry suggests that technological innovation displaying interdependencies among the parts is greatly facilitated by common ownership of the parts.

The above analysis has concerned itself with the organizational mechanics of getting an innovation commercialized. The post-commercialization market performance of the innovator is also a very significant matter, which has been dealt with elsewhere (Teece, 1986). Suffice to say that ownership by the innovator of the supporting assets and skills needed to ensure competitive supply of the new product, or of existing products based on the new process, is often required to ensure that the rent stream from the innovation is shielded from capture by 'fast seconds'. The exceptions are where the appropriability regime—that is, the protection afforded the new product or process by patents, copyrights, trade secrets and inherent 'hard-to-copy' aspects of the innovation—is extremely tight. When property rights are difficult to establish and where imitation, either through 'inventing around the patent' or reverse engineering or other activities is relatively easy (i.e. the appropriability regime is weak), then the innovator needs to own or otherwise control the relevant cospecialized assets to be able to impede the imitator's efforts to take the product/service to market or more advantageous terms than the innovator (Teece, 1986). Since cospecialized assets in marketing, distribution and manufacturing are often aligned vertically, vertical integration may be required in order to assist the innovator in capturing the rent stream generated by the innovation.

This analysis is consistent in part with an alternative argument which has been made. Several writers, including Utterback (1978), have speculated that older, vertically integrated firms will have a greater commitment to old technology because of the large technology-specific investments they have made upstream and downstream. The phenomenon to which Utterback refers is simply that innovators may resist cannibalizing the value of their own irreversible investments. Put differently, innovating firms, integrated and otherwise, that have laid down innovation-specific investments will generally not be the first to commercialize new innovations which will

impair the value of existing assets. They will do so only under competitive threat, or if by doing so the present value of the profits from the new innovation will outweigh the losses from the old.

Since vertically integrated firms often have specialized investments in place and since a primary rationale for vertical integration is to protect specialized investments from recontracting hazards, it is to be expected that vertically integrated firms will have a higher proportion of their asset base which is dedicated to particular technologies than do non-vertically integrated firms. Thus a monopolist which is vertically integrated and has assets specialized to the old technology may indeed delay the commercialization of new technology if it is confident that it does not face competitive threats. Were such delay to occur, it is caused not by vertical integration as such, but by the fact that the innovator, by assumption, owns assets dedicated to the old technology—assets whose value will be impaired by the new technology. Hence, it is theoretically possible that vertically integrated firms, because they own assets dedicated to the old technology, may retard the commercialization of new technology when the following conditions hold: the vertically integrated firm is the innovator; the vertically integrated firm is a monopolist or possibly a colluding oligopolist; the vertically integrated firm has made investments in the old technology which will be impaired in value by introduction of the new; the innovation destroys a rent stream with a net present value to the innovator greater than the rent stream that it would serve to create.

Such instances are likely to be infrequent and, moreover, the impairment to the commercialization of innovation flows fundamentally from a combination of market power in the presence of sunk costs. Whether the dampening of commercialization is socially as well as privately desirable will swing on the magnitude of the externalities generated by the new technology relative to the old.

Transacting for know-how across enterprise boundaries: collaborative arrangements and technological change

The analysis so far has examined the properties and stressed the virtues of organizational arrangements in which research and development proceeds as an in-house activity. As indicated at the outset, however, contract research is in some cases viable. Moreover, once technology is produced, in many cases it can be traded (licensed) in the market for know-how which is increasingly international in its scope (Teece, 1981).

Indeed, the vertically integrated enterprise with in-house research and its own manufacturing, distribution and sales is in some industries being joined by almost pure research enterprises. Relatedly, incumbent firms are increasingly engaged in extensive collaborative dealings with other firms, especially research-oriented, new business enterprises. In this section, various reasons for these developments are explored and the implications for the organization of research assessed.

The organizational arguments (section on 'The integration of R & D with production') favoring in-house research in order to avoid contractual difficulties rests on a fundamental assumption, namely that the firm has the inventive capacity to develop competitive technology in-house. However, given that the institutional loci of new technology in the US are diverse and include the universities, other not-for-profit institutions, and government laboratories, there is a high probability that from time to time established firms will have to source technology externally. When knowledge accumulation is cumulative, then established enterprises can generally build upon existing competences in order to develop new technologies in a timely and cost-effective fashion. Occasionally, fundamental breakthroughs in science and technology occur which do not build upon incumbent firms' competences. Such developments constitute what was referred to earlier as paradigm shifts. When these shifts cause the institutional locus of innovation to lie external to incumbent firms and the new knowledge in question is proprietary and difficult to copy, then the opportunity for licensing and other forms of collaboration become manifest. This appears to be the case with biotechnology where the key breakthroughs have been generated within the universities, and this in turn has spawned several hundred, small biotechnology firms usually founded by scientists.[6] Incumbent pharmaceutical firms, such as Eli Lilly, Merck, and Johnson & Johnson, have seen both the opportunity and the threat posed by the new biotechnology[7] and the new enterprises that have been spawned to develop it.

The opportunity stems from the ability to develop and commercialize new products which will open up new markets while employing at least part of the incumbent firms' fixed costs in plant, equipment, distribution and human capital. The threat stems from the possibility that the new technology will render obsolete incumbent firms' products, facilities and capabilities. Research collaboration (such as in licensing, joint R & D) is attractive to the incumbents for those reasons and more; to the new business firms it is often a source of capital. It also can provide access to downstream assets, particularly marketing channels. In short, collaborative research can occur alongside in-house research in order to bolster the technological capabilities of incumbents and in order to enable new business firms—which may begin as stand-alone research ventures—to continue to fund research and to acquire the ability to integrate vertically into manufacturing, marketing and distribution. Needless to say, the market for know-how is likely to encounter many of the contractual difficulties described in the section on 'The integration of R & D with production' in this chapter; as the biotechnology industry evolves, it is to be expected that the new business firms will take on a more classical structure. Indeed, the two leading biotechnology firms, Genentech and Cetus, are actively pursuing a vertical-integration strategy with respect to their key businesses.

A shift in technological paradigm is not the only reason why in-house research gets displaced as the main driver of a firm's technological capability. Scale issues, as with the development of new jet engines or large central office telecommunication switches, may require collaboration. So may pure incentive considerations, as when a salaried scientific position is no match for the outcome-driven incentives of the new business firms.[8] But the pervasiveness and durability of in-house research is worthy of comment. The 1970s and 1980s have certainly exhibited important changes in the way that research is organized, but even the new enterprises seem rapidly to adopt structures which eschew contract research, except in the very early stages of industry development. The robust nature of the organization of research in the modern corporation would thus appear to be apparent, even though the corporation is taking on certain 'post-modern' features.

Implications and conclusions

The above analysis of the firm and its relationship to technological change helps us understand not only the nature of the firm, but also sheds light on some topical managerial and public policy issues raised in the introduction. Some of these implications are briefly summarized below.

Stand-alone laboratories and hollow corporations

The natural organizational home for research appears to be inside the corporation, alongside production/operations. This model seems to be dominant for large corporations, and to a lesser extent for smaller corporations, as it facilitates interaction between the users and providers of new technology. It also avoids the difficulties associated with writing, executing and enforcing R & D contracts. Relatedly, managerial decisions to 'hollow out' the corporation by out-sourcing components and other subsystems may have the indirect effect of impairing the innovation process by establishing barriers to the transfer of information between research and manufacturing, possibly causing future designs to be less sensitive to manufacturability concerns. It may also serve to enhance the capabilities of competitors and potential competitors. Out-sourcing runs the risk of creating circumstances whereby innovators are no longer able to profit from innovation, despite the fact that they are highly innovative.

Diversification economies

The evolution of technology is often driven by certain technological imperatives which induce firms to gravitate in certain technological directions. This technological drift is often highly constrained, causing firms to articulate focused competences ('core' businesses). Sometimes scope economies become available when a key internal competence affords multiple application. Diversification is often a desirable organizational response for a set of reasons similar to why research is better supported in-house rather than via contracts. This suggests that the core business of an enterprise typically has a technological underpinning, and that efficient diversification is likely to be driven by technological imperatives. Hence the focused or laterally diversified enterprise—where corporate diversification tracks underlying technological imperatives—is likely to be a characteristic of economies in which efficiency concerns drive diversification decisions. Unless tax and technology transfer issues are visible drivers of corporate diversification decisions, one may be entitled to suspect that corporate diversification is driven by factors not consistent with stockholder, wealth-seeking behavior.

Vertical integration and technological innovation

The analysis presented in this section indicates that when a stream of innovations has significant systems ramifications, then vertical integration is likely to facilitate the commercialization of an innovation, if not its initial development. Of course, not all technology possesses strong systems interdependencies, and many that do involve interdependent organizations which do not afford vertical integration opportunities.[9] However, when the relevant organizational domain is within the range of feasible integration, vertical integration is likely to facilitate innovation, and may well be required if a stream of systematic innovations is to be commercialized in a timely fashion. Vertically integrated firms may also thwart the introduction of new technologies when the innovation would have the effect of destroying the value of investments in place, and the vertically integrated developer of the innovation is not threatened by a competitive technology.

New business firms and research collaboration

The analysis earlier in this chapter suggested the importance of performing research in-house. Contract research is usually but not always a poor substitute. However, opportunities for many other forms of collaboration, such as R & D joint ventures, do exist. Indeed, they may represent an imperative in instances where the firm contemplating conducting research lacks the desired skills and is unable to acquire them in the labor market. This might actually characterize even research-intensive firms when a shift in the technological paradigm renders the existing skill base of the enterprise obsolete or irrelevant.

Collaboration between established firms and universities, and between established firms and new business firms that possess the relevant skills may therefore be necessary. The incumbent firms are likely to possess marketing and manufacturing assets of great value to a new business firm, while the new business firm may have research findings, capabilities, or possibly even products of great value to the established firms. Circumstances such as these provide opportunities for collaboration.

Collaboration by definition falls into neither the 'contract research'[10] nor the 'in-house research' categories identified earlier. Its ubiquity in no sense destroys the argument made earlier in favor of in-house research. The

presumption in favor of in-house research can be readily overturned, for a transitory period, when the sources of know-how lie external to the firm and cannot easily be acquired through 'hiring in' technical and scientific personnel. In these instances, co-development activities and R & D joint ventures may make good sense. Often collaboration in research and development is part of a larger arrangement involving production and marketing. Clearly, a more complete understanding of the organization of research in a capitalist economy requires our assessment of a broader set of institutions, including universities, that condition the environment in which technological change proceeds.

Notes

1. Edison's Menlo Park laboratory, which exmployed sixty-four people by February 1880, has been called the world's first industrial research laboratory, but it was not a prototype of those to follow. It was organized to give vent to the creative genius of one man only—Thomas Edison (Friedel and Israel, 1986).
2. Specifically, Teece and Armour (1977, pp. 56–7) noted: 'Universal Oil Products—Houdry is a similar example; Scientific Design would be an analogous example in the chemical industry—is an engineering, design, and research company that is not integrated into production, refining, transportation, or marketing, and yet has made important contributions to technological innovation in the petroleum industry. The rather narrow research activities of Universal Oil Products should, however, be indicated here, lest it be assumed that this example could represent an appropriate model for the entire industry. First, it would seem that where the research objectives are simple and obvious, a nonintegrated research and development firm like Universal Oil Products may not be particularly disadvantaged. For instance, the development of higher octane gasolines, or the development of processes to meet new environmental standards, do not involve the formulation of complex research objectives. Universal Oil Products' research seems to have been confined to meeting simple objectives relating to the refining function: the company has not been responsible for innovations on lubricants, petrochemicals, or exploration and production. It mainly performs applied research and avoids high risk endeavors. The nonintegrated research and development firm, though performing useful services for the industry and consumers, does not seem a sufficiently robust organization to absorb the full gamut of current research and development activities in the industry. The specialization that has emerged seems to have advantages, and an assumption that a nonintegrated research and development structure could successfully perform the current industry portfolio of research and development projects would not seem to be grounded on an understanding of the many subtleties of the research and develoment process.'
3. Richard Tybout (1956) succinctly characterizes the cost-plus contract as follows: 'The cost-plus fixed fee contract is the administrative contract par excellence. For the market mechanism, it substitutes the administrative mechanism. For the profit share of private entrepreneurs, it substitutes the

fixed fee, a payment in lieu of profits forgone. And for the independent private business unit, it substitutes the integrated hierarchical structure of an organization composed of an agency . . . and its contractors' (1956, p. 175).

4. 'Consolidated Edison Company of New York: the development of an atomic power plant', Weapons Acquisition Project, Harvard Business School, December 1959.
5. Despite the fact that the ultimate objective of R & D programs is to produce innovations, not simply to dissipate resources on R & D activities, expenditure data can be viewed as a useful proxy for innovative performance in that it reveals the intensity of innovative activity. Furthermore, if the discount rate facing non-integrated firms is similar to that facing integrated firms and if similar risk preferences exist across the management of these firms, the higher productivity per dollar of research expenditure posited in vertically integrated firms implies that, *ceteris paribus*, such firms will devote more resources to R & D.
6. Over 400 by 1987 in the US alone.
7. The new biotechnology consists of three general techniques: recombinant DNA (rDNA), cell fusion (monoclonal antibody technology), and the novel bioprocessing technology.
8. Williamson (1985) traces these differences to the 'low powered' incentives of large organizations versus the 'high powered' incentives of smaller entrepreneurial firms.
9. See Horwitch and Prahalad (1981) for a discussion of the variety of institutions that are most often involved in the innovation process.
10. Note that contract research is different from licensing. Patent licensing, for instance, involves the sale of intellectual property rights, usually subject to certain restrictions. General licensing involves the sale of scientific and technological assets already developed. Contract research involves the development, for fee, of technological assets.

References

Arditti, F. D. (1968), 'On the separation of production from the developer', *Journal of Business*, vol. 41, no. 3, pp. 317–32.

Brown, G. S., Davidson, F. P. and Little, H. F. (1976), 'The systems perspective: concepts related to sustained efficiency of the United States telecommunications network', F.C.C. Docket 20003, Bell Exhibit 62.

Dosi, G. (1982), 'Technological paradigms and technological trajectories', *Research Policy*, vol. 11, pp. 147–62.

Frankel, M. (1955), 'Obsolence and technological change in a maturing economy', *American Economic Review*.

Friedel, R. and Israel, P. (1986), *Edison's Electric Light*, New Brunswick, NJ, Rutgers University Press.

Goldberg, V. (1977), 'Competitive bidding and the production of accountant information', *Bell Journal of Economics*.

Horwitch, M. and Prahalad, C. K. (1981), 'Managing multi-organization enterprises: the emerging strategic frontier', *Sloan Management Review*, Winter.

Kindleberger, C. P. (1964), *Economic Growth in France and Britain, 1851–1950*, Cambridge, Mass., Harvard University Press.

Malmgren, H. (1961), 'Information, expectations, and the theory of the firm', *Quarterly Journal of Economics*, vol. 75, August, pp. 399–421.

Marx, T. (1976), 'Vertical integration in the diesel-electric locomotive building industry: a study in market failures', *Nebraska Journal of Agricultural Economics*, vol. 15, no. 4, Autumn, pp. 37–51.

Miller, R. and Sawyes, D. (1970), *The Technical Development of Modern Aviation*, New York, Praeger.

Mowery, D. C. (1983), 'The relationship between intrafirm and contractual forms of industrial research in American manufacturing, 1900–1940', *Explorations in Economic History*, vol. 20, pp. 351–74.

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

Roberts, Edward (1979), 'Stimulating technological innovation: organizational approaches', *Research Management*, November, pp. 26–30.

Rosenberg, N. (1969), 'The direction of technological change: inducement mechanisms and focusing devices', *Economic Development and Cultural Change*, vol. 18, pp. 1–24.

Scherer, F. M. (1965), 'Size of firm, oligopoly and research: a comment', *Canadian Journal of Economics and Political Science*, May, pp. 256–66.

—— (1980), *Industrial Market Structure and Economic Performance*, Chicago, Rand McNally.

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

Sloan, A. P. (1964), *My Years with General Motors*, New York, Doubleday.

Stigler, G. J. (1956), 'Industrial organization and economic progress', in L. D. White (ed.), *The State of the Social Sciences*, Chicago, University of Chicago Press.

Teece, D. J. (1976), *Vertical Integration and Vertical Divestiture in the U.S. Petroleum Industry*, Stanford, Institute for Energy Studies.

—— (1977), 'Technology transfer by multinational firms: the resource cost of international technology transfer', *Economic Journal*, June.

—— (1980), 'Economies of scope and the scope of the enterprise', *Journal of Economic Behavior and Organizations*, vol. 1, no. 3.

—— (1981), 'The market for know how and the efficient international transfer of technology', *The Annals of the Academy of Political Social Science*, November.

—— (1983), 'Towards an economic theory of the multiproduct firm', *Journal of Economic Behavior and Organizations*, vol. 3.

—— (1986), 'Profiting from technological innovation', *Research Policy*, December.

Teece, D. J. and Armour, H. O. (1977), 'Innovation and divestitute in the U.S. oil industry', in D. J. teece (ed.), *R & D in Energy Implications of Petroleum Industry Coorganization*, Stanford, Stanford University Institute for Energy Studies.

Thompson, J. D. (1967), *Organizations in Action*, New York, McGraw-Hill.

Tybout, R. (1956), *Government Contracting in Atomic Energy*, Ann Arbor, University of Michigan Press.

Utterback, J. M. (1978), 'Management of technology', in A. Hax (ed.), *Studies in Operations Management*, Amsterdam, North Holland, pp. 137–60.

White, L. (1971), *The Automobile Industry since 1945*, Cambridge, Mass., Harvard University Press.

Williamson, O. E. (1975), *Markets and Hierarchies*, New York, Free Press.

—— (1985), *The Economic Institutions of Capitalism*, New York, Free Press.

13 The R & D function: corporate strategy and structure

Neil Kay

Department of Economics, Herriot-Watt University, Edinburgh

R & D in the firm: function, strategy and structure

The purpose of this chapter is to examine the R & D function, strategic problems, internal organisation, and the interrelationship between them, from the perspective of economics. We shall analyse the economics of R & D activity in terms of four basic features or characteristics, and then consider the implications of these characteristics for function, strategy and structure. We hope to show that these characteristics are recurring features of analysis, often underlying a wide variety of issues and problems that have frequently been treated independently of each other in separate literatures and approaches. We start by examining the four characteristics, before examining in turn their implications for R & D activity, strategy and internal organisation.

The characteristics of R & D activity

The four characteristics or features that have central importance for the economies of R & D acxtivity are non-specificities, lags, uncertainty and costliness. *Non-specificity* in this context is relevant at the level of the product and the firm. Much R & D is not product-specific in so far as a particular piece of work may feed into a variety of end products, the R & D generating technological synergies, or economies of scope. Also, much R & D activity is not firm-specific, generating externality and property right problems. Both questions are likely to be very important for the firm; low product-specificity may allow the firm to spread R & D costs over a variety of product lines, while low firm specificity may signal a weak or low competitive advantage for the firm in its R & D activity. *Lags* and delays are a typical feature of R & D activity, a given piece of R & D often taking many years before it is embodied in commercial ventures, if at all. In themselves, lags are not necessarily an intractable problem, but they may directly contribute to other problems such as dangers of losing proprietary knowledge (low firm specificity), cost and uncertainty. *Uncertainty* is also a pervasive problem, uncertainty in this context meaning unmeasurable or non-insurable uncertainty (Knight, 1921), in contrast to predictable or measurable risks of an actuarial nature. Uncertainty here can be classified into general business uncertainty, which refers to all decisions concerning

the future; technical uncertainty, which is concerned with achievement of specified performance and cost level; and market uncertainty, which refers to the possible achievement of a commercially viable product or process. R & D work can be faced with problems of uncertainty of all three types (Freeman, 1982, Chapter 7). *Cost* levels and associated resource commitment also tend to pose problems, though this can vary from sector to sector. The barriers to entry, or even to continuance, posed by high and/or increasing R & D economies of scale and scope have become a major issue in some sectors like aerospace and automobiles. Just as lags may not pose insuperable problems in certain circumstances, so also cost level itself should not be a problem if there are no significant problems of knowledge and information in the market-place. Such issues do become important, however, if R & D cost levels exceed the internal financing capability of the firm, and there are information barriers to external capital market financing of corporate projects.

The impact of these factors generally varies as a project moves from basic research through applied research into development and then final introduction or innovation. Generally, non-specificities, lags and uncertainty tend to decrease, while cost levels and associated resource demands tend to increase, as a project moves downstream through the various stages towards final innovation. As far as non-specificities are concerned, both product- and firm-specificity of R & D tends to increase as a project moves towards final launch. For example, in laser R & D, basic research may be ultimately applicable to a wide range of applications in laser technology, applied research is likely to be concerned with a narrower range of potential applications, say in measuring devices, while resulting development work is liable to be specific to a specific measuring device or highly related group of devices. This tendency for product specificity to rise as a project moves through the various R & D stages may also create parallel tendencies for firm specificity to rise in the same direction; the extent of external applications due to leakage of technical information is likely to be directly related to product non-specificity. Thus, externalities may be more important for earlier, upstream research activity, especially basic research. Any tendency for firm-specificity to vary in this fashion may be reinforced to the extent that development work reflects tacit knowledge that may not easily diffuse externally, and also to the extent that basic research involves appropriability problems such as inapplicability of patent protection.

The other features tend to vary in a more obvious fashion as projects move along basic research, applied research, development, to introduction. Lags to final innovation will tend to be cumulative through the various stages and will tend to shorten towards and through development (Kay, 1979, pp. 23–4). Uncertainties tend to increase the further a stage is removed from final innovation (Freeman, 1982, p. 150), and so degree of uncertainty in its various manifestations is likely to diminish as a project moves through its various stages to completion. Finally, the cost of R & D activity tends to increase as projects move from earlier stages through to development and from laboratory experimentation to prototypes and pilot plants (Mansfield, 1968, p. 78; Schon, 1976, pp. 40–2).

In the next section we shall examine how these characteristics individually and collectively contribute to important issues and problems in R & D management and behaviour.

The four characteristics and R & D behaviour

It follows therefore that non-specificity, lags, uncertainty and costliness are each common features of R & D, though the relative importance of respective characteristics may vary with technology, firm or even time period. The extent and significance of the first three characteristics will tend to diminish as projects move from earlier upstream stages towards eventual innovation, while costliness of projects and associated resource commitments frequently increase in the same direction. Any reasonably innovative project is likely to encounter issues of non-specificity, lags, uncertainty and cost that could have important implications for their competitive position in the market-place. For example, are product specificities low enough to provide synergies and spread R & D costs? Are firm specificities strong enough to avoid appropriability problems? Are lags short enough to facilitate first-mover advantages? Are uncertainties sufficiently controllable to guide resource direction and reassure the capital market? Will cost considerations be low enough to permit internal financing?

Not only will questions like these vary in importance from project to project; they will vary in importance as projects and related or derived projects move downstream towards final completion. In this section we explore the implications of these general issues for a number of problems in R & D behaviour.

The financing of R & D and the importance of uncertainty

Uncertainty is a dominant characteristic influencing the financing of R & D both at project level and at the level of the R & D function overall. To start with, very little R & D work is financed by the external capital market, most being internally financed (Freeman, 1982, p. 149). However, this may conceal a greater level of capital market response to R & D activity than is apparent at first sight, since the external capital market may be strongly influenced in their overall level of support for the company by general R & D performance, as well as signals relating to proposed activity.

One element that may impede efficient linking of external capital and internal R & D is possible conflict of interest in information disclosure as far as capital market and product market is concerned. Improving the quality and detail of R & D plans available to the capital market may have a detrimental effect on a firm's competitive advantage in the product market. These conflicts may constitute limiting factors on the potential efficiency of

external financing, and the problems may be exacerbated to the extent that corporate actors are liable to indulge in opportunistic behaviour and misrepresentation. If, despite these problems, external financing is undertaken, say, by government agencies, uncertainty creates problems in contract design. A cost-plus system would leave the sponsoring agency vulnerable to moral hazard and opportunism, while fixed-price systems might make it difficult to find R & D-conducting firms willing to bear the uncertainties and associated costs (see Chapter 12).

The frequently observed optimistic bias in estimating R & D costs and lag times (Freeman, 1982, pp. 151–6) is also attributable to uncertainty, either because no allowance is made for uncertainties, 'bugs' and surprises, or because project estimators may opportunistically abuse an informationally superior position to gain project approval by deliberately underestimating costs and lags.

Uncertainty also creates time-cost trade-offs; if a target is required urgently, as in cancer research, many tasks may have to be carried out simultaneously, increasing the chances of duplicated learning, dead-ends and diminishing returns (Mansfield, 1968, p. 72; Freeman, 1982, p. 151). If the target can be approached more slowly, many R & D tasks can be carried out sequentially, permitting transference of learning and experience, with consequent economising on resources. Finally, the same problems of uncertainty encountered in this problem area also impede construction of R & D budgets in a rational, aggregative, bottom-up fashion. As a consequence, most large firms allocate annual funds to the R & D function on a rule-of-thumb basis such as percentage of sales (Mansfield, 1968, p. 62; Kay, 1979, pp. 72–7). The actual budget rule often evolves through decision-makers learning what is the 'appropriate' budget for their firm (Freeman, 1982, p. 163).

Basic research: the extreme case

The further upstream a project is located in the basic research–applied research–development–introduction progression, the more lags, uncertainties and non-specificities assume importance in the resource allocation process. This is especially the case for basic research activity. The lags and uncertainties involved may discourage private investors (Freeman, 1982, p. 168) and these problems are likely to be compounded by the existence of non-specificities in the form of externalities. Consequently, government intervention or support for basic research is likely as a result of these market failure problems, though government support for basic research has traditionally been biased towards support for universities rather than corporations. For those firms that do conduct basic research, the diversified firm is likely to have an advantage, since a broad portfolio of businesses means that the various unpredictable and unexpected results are more likely to be internalised within corporate businesses (Nelson, 1959).[1] The 3M Corporation is an example of a highly diversified firm which has a successful track record of exploiting radical, innovative opportunities in this way, often within divisions unrelated to those that developed the original idea.

Demand-pull[2] theories of innovation are likely to be less relevant the further upstream a project is located. If demand-pull theories are relevant at all, it is likely to be in the development stage when work is close to completion, less uncertain, and more specific and precise in its output; Freeman (1982, p. 103) points out the pull of the market operated as a complementary force to technological momentum in many cases where the market demand was *urgent* and *specific*. The further back towards basic research a project is located, the more supply-side science and technology-push arguments are likely to be relevant.[3]

Winners and losers

Being first to introduce a new product or process does not guarantee success, and indeed the four characteristics of R & D may combine against the first-in. The first-in may incur severe problems of uncertainty, delay and cost, while non-specificities may contribute to rapid leakages of technical knowledge externalities to potential competitors. Mansfield (1985) produces evidence to show that information on technical developments typically leaks out very rapidly to competitors in a wide range of technologies. The second-in may exploit such non-specificities to cut down on the uncertainties, delays and costs incurred by the pioneer. The pioneer may have first-mover advantages, but these factors may erode them partially or totally. Sperry's loss of an early lead in commercial computer development to IBM was a major example of this type.

Consequently, the link between a firm's own R & D and its subsequent growth is highly tenuous at best, though for a given industry as a whole there is typically a stronger and observable relationship between industries R & D and industry growth (Freeman, 1982, p. 164). The instabilities associated with firm level tend to smooth out at industry level, while many of the externalities will work themselves out within industry boundaries, strengthening the link between industry R & D and growth.

Implications of possible recent changes in the characteristics

The cost factor has become an even more important influence in some sectors in recent years. Previously, merger and takeover represented strategic devices for spreading escalating R & D costs and exploiting internal economies. In some sectors, such as aircraft and automobiles, mergers and takeovers may have reached saturation point for nationalistic and anti-trust reasons. Joint ventures and licensing agreements have grown in importance in recent years and, in at least some cases, represent attempts to spread R & D costs in cases where merger is not feasible or desirable. R & D consortia or clubs have also evolved for similar reasons in some sectors such as electronics.

The chapter by Teece discusses joint ventures in more detail, but there are also issues of relevance to the analysis of this chapter. The creation of such R & D consortia or clubs may provide cooperative gains for the participating members in the form of technological advantages that would have been difficult or too costly for individual members to develop alone. We would expect such cooperative ventures to emphasise more upstream non-product-specific activity, since research of this nature is more likely to benefit the group as a whole, or a significant proportion of its members. More downstream product-specific development activity might lead to possible direct competitive conflict between group members and could have a zero-sum quality for participants in the venture. Consequently, Nelson (1984) suggests that such ventures are more likely to be appropriate for the exploitation of generic research programmes applicable to a variety of subsequent development programmes. Some evidence of this is provided by Peck (1986), who cites the example of MCC, a private micro-electronics and computer technology joint R & D project involving twenty-one US companies. Its research programmes are generally consistent with Nelson's definition of generic research; Peck gives as an example the VLSI/CAD programme in MCC which does not design specific circuits but instead seeks to develop methods of computer designing circuits (p. 220).

Another possible change that may have significance for the conduct of research is that the perceived lag between conduct of basic research and eventual commercial application may have shortened in some cases, such as certain areas of biotechnology. Scientific norms of openness and active dissemination of research results may be compromised if commercial applications are expected in the relatively near future; the compartmentalisation of the historically distinct traditions and norms of science and commerce has traditionally been facilitated by the existence of buffers in the form of long lags weakening the profit implications of basic research for individual researchers.

Therefore, the four characteristics, and changes in these characteristics, have fundamental implications for the conduct of R & D activity. In the next section we shall see that these same characteristics have similarly profound implications for corporate strategy and internal organisation.

Strategy and structure implications

In this section we shall look at some implications of the four characteristics for the strategy and structure of individual corporations. We shall devote more attention to problems of structure or internal organisation, since many of the problems relating to strategy have already been introduced and discussed in the previous section. Freeman (1982) provides a synopsis of strategy types that incorporate or reflect a number of problems discussed in the last section. As far as analysis of strategy itself is concerned, Freeman (1982) produces a useful basis for analysing different types of R & D strategy. The *offensive* strategy will be appropriate if there are particular advantages to being first in with a particular innovation. Here protection of property rights (especially non-specificities) and lags required for competitive response are critical elements in deciding whether or not to adopt an offensive strategy. The *defensive* strategy is still likely to involve a high level of R & D, but the firm is prepared to react and follow offensive innovators, possibly with some degree of product differentiation. Obviously, if an offensive innovator finds that non-specificities benefit other firms in the form of externalities, this facilitates defensive strategies, and, as we saw in the previous section, such circumstances may be the norm rather than the exception. The ability to respond quickly and reduce lags is also important. As Freeman points out (1982, p. 178), a science-based firm's R & D strategy may contain mixtures of offensive and defensive strategies. IBM is an example of a company which has successfully and fairly consistently pursued a defensive strategy, mobilising considerable technical and marketing resources to respond to external technological threats. The *imitative* strategy does not attempt to match the offensive and defensive innovators in terms of skills and is prepared to follow some way behind if it enjoys particular advantages in terms of cost, tariffs or supplies. The *dependent* strategy is usually followed by smaller firms with subordinate subcontracting roles in which they do not initiate new products but accept specifications and conditions imposed by dominant firms. Component manufacturers in the automobile industry have typically adopted such a role, though there are interesting signs that even in this sector the threat from Japanese manufacturers is likely to lead to merger and consolidation amongst parts-makers. Amalgamation would help create the critical mass necessary to take a more active and leading role in component and sub-system innovation (*The Economist*, 23 May 1987, p. 80). The *traditional* strategy is based on absence of technological innovation in a market which is benign and slow changing, while the *opportunist* strategy is based on entrepreneurial perception of niches that may not require substantial in-house R & D.

The four characteristics discussed earlier may all influence the possible strategies a firm will choose. For example, cost, uncertainty and lags may dissuade smaller, specialised firms from adopting offensive or defensive strategies, while non-specificities and second-in advantages discussed earlier may persuade firms to adopt a defensive rather than an offensive strategy. On the other hand, if it is possible to disengage the relatively cheaper upstream inventive stages from the more expensive downstream development work, small firms may have a comparative advantage on these cheaper, earlier stages. Large firms may be better placed to indulge in costly full-scale development (Freeman, 1982, p. 137; Williamson, 1975, p. 142). The role of structure or internal organisation in R & D activity is also influenced by the four characteristics. Williamson (1975), building on Chandler (1962), provides an analysis that will help introduce basic ideas which we can then develop further using our earlier analysis of the four basic characteristics.

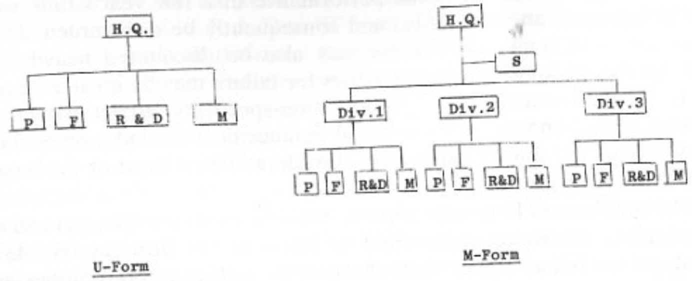

Figure 13.1 Basic hierarchies

Williamson contrasts the functional, or U-form, structure which is more appropriate for smaller, specialised firms with the multi-divisional, or M-form, structure which is likely to evolve in large diversified firms. In a specialised firm, the similarities and synergies between groups of products are so strong that functional specialists are liable to have to spend considerable time talking to each other and coordinating plans and schedules. Consequently, it is logical to group them together in a functional home as in the U-form example in Figure 13.1.

The M-form structure will tend to be more appropriate for a large, diversified firm. Divisionalisation creates natural decision units in the diversified corporation, putting together those functions responsible for a product or group of products. The divisions are assigned responsibility for operating and short-term strategic decisions, so as to reduce the number of levels in the hierarchy that have to be crossed before an inter-functional decision is arrived at, reducing delay, loss and distortion of information, and permitting top-level management to concentrate on long-run, strategic decision-making. Further, a competitive internal capital market may be created, comparability of divisions being facilitated by the existence of a uniform project yardstick for performance at middle (divisional) levels.

Therefore, the M-form creates a rational structure for the management of large, diversified firms. The problem that R & D poses for this solution is that of all the functions it is the least likely to be amenable to such treatment. We can see why by comparing the characteristics of reasonably innovative R & D work with the characteristics of economic activity appropriate to divisional operations. On the four characteristics of time, uncertainty, non-specificity and cost, reasonably innovative R & D work scores badly on the first three counts as far as divisional implications are concerned. Profit-centre operation tends to be short run; divisional managers are not only commonly assessed on the basis of annual performance, they are typically highly qualified and mobile general managers for whom divisional performance in a few year's time would often represent an externality, and consequently be disregarded. Highly uncertain and long-term returns may also be discounted heavily by a managerial set for whom the penalties for failure may be greater than the rewards for success.[4] The product non-specificity aspect would also reinforce the tendency for divisional management to under-invest in, or neglect, R & D due to externality considerations, if most of the benefits accrue to other profit centres.

The problems of long-time horizon, non-specificity (or synergy) and high uncertainty are more appropriate to levels in the firm responsible for strategic overviews rather than short-term, specific responsibilities as in divisions. The further we move upstream towards basic research activity, the more these problems become exacerbated for divisional management. Therefore, the tendencies towards centralisation of R & D are accentuated the further upstream the R & D is located. This is indicated at 'S' level for the M-form in Figure 13.1. The fourth characteristic, cost, may work against the inclusion of even development work in some divisions if the system under development is extremely costly and complex. Divisions often have extremely limited discretionary access to internal funds, and costly development plans may exceed their discretionary limits. Accordingly, financing decisions, or even the development work itself, may have to be shoved further up the corporate hierarchy.

In those circumstances, it is reasonable to ask why *any* R & D should be conducted at divisional level. Low product-specificity, long lags, extreme uncertainty and high cost all provide barriers to the effective divisionalisation of R & D. If even one of these characteristics is present in a particular R & D project, it impedes divisionalising the project and provides pressure for incorporation at 'S' level in Figure 13.1. Further, even if some projects do not possess any of these characteristics to a significant extent, there may still be incentives to allocate them to 'S' level to avoid diseconomies from splitting the R & D function.

There are, in fact, counteracting pressures to leave some or all R & D at divisional level. Firstly, removing a major function such as R & D from divisional responsibility impairs the internal capital market in so far as R & D cost is now shared between divisions, reducing the extent to which divisions can be treated as independent profit centres. Secondly, in a major study of factors influencing success or failure of innovation,[5] the factor which discriminated most clearly between success or failure was whether or not users' needs were understood (Freeman, 1982, p. 124). Separating R & D from divisional marketing could inhibit the integration of technological possibilities and designs with consumer requirements. Therefore, the location of R & D in the corporate hierarchy is likely to be a complex problem in practice, involving trade-offs between divisionalisation/centralisation advantages and disadvantages. Different companies evolve different solutions to these conflicts. At a relatively early stage in its post-

war development. General Electric discovered that its divisions were not innovating largely for the reasons discussed earlier, and responded by reallocating much of its R & D to 'S' level. Du Pont's solution to similar problems was to split its divisions' budgets into two components, one for operations and one for innovations, and monitor the respective components separately.

In fact, the formal hierarchical designs discussed above face potentially severe limitations imposed by the nature of R & D and product life-cycle considerations. Hierarchical bureaucracies of the type described above fall into the category of mechanistic systems as defined by Burns and Stalker (1968).[6] *Mechanistic* systems are characterised by functional specialisation, precise roles, vertical interaction between managers and formal hierarchical relationships. This form was identified by Burns and Stalker as being appropriate to technologically stable conditions. *Organic* systems are characterised by informal lateral relationships, networks rather than hierarchies, continual redefinition of tasks, and broadly specified responsibilities. These characterisations are ideal types and in practice organisations operate along a continuum on which these descriptions represent polar extremes.

Mechanistic systems encounter problems in incorporating innovative decisions on their agenda, and are likely to face extreme difficulties in rapidly changing environments. Since innovation constitutes a break in existing standard operating procedures and programmes, mechanistic structures typically have great difficulty in accommodating relevant decisions within existing routines.

Relevant information is liable to be ignored or mistreated because it does not fit into existing classifications or may face delays in being acted on as it is referred up the hierarchy. Even if the system is set up to signal significant data on innovation, in a highly innovative environment this would lead to the senior management fire brigade facing a number of alarm bells ringing simultaneously. Innovation, especially in turbulent environments, is characterised by non-programmable, surprising, routine-breaking information, and the mechanistic structure typically encounters severe difficulties in this area of decision-making. The characteristics of R & D uncertainty and non-specificity discussed earlier are particularly relevant here; Burns and Stalker point out that the mechanistic system is designed to deal with tasks that are *precise* and *specific*; uncertain, surprising tasks with a high degree of non-specificity will be ignored or mishandled by such a system.

The organic system, with its absence of pre-set rules, roles and responsibilities, is better equipped to facilitate innovativeness in rapidly changing environments. The form sacrifices the possibilities of static economies from functional specialisation and division of labour, but this can be a cheap sacrifice in conditions of rapid technological change, since these economies may not be obtainable in any case. What it provides instead is flexibility and responsiveness; the non-specificity and uncertainty inherent in innovation finds parallels in the non-specificity and uncertainty typically surrounding relationships and roles in the organic system.

The principles differentiating organic from mechanistic systems in Burns and Stalker's early analysis have been embodied in a variety of organisational designs that have evolved in recent years. For example, project teams or task forces with limited life spans may be set up to deal with innovative opportunities, cutting across formal organisational hierarchies and dealing with non-specificities by focusing on 'the innovation' as the unifying concept. Matrix management is a more elaborate and complex solution displaying both mechanistic and organic features and has been adopted by firms facing turbulent environments, including ITT, Monsanto, ICI and Lockheed. In a matrix structure an individual typically has simultaneous responsibilities to a functional home (e.g. R & D, production, marketing) and to a specific project. In principle, the functional line of responsibility provides mechanistic static economies from functional grouping and specialisation, while the project line of responsibility provides organic dynamic efficiency gains by focusing and integrating at the level of the particular project or innovation. In practice, dual responsibilities and confusion of responsibilities may inhibit the extent to which firms operating a matrix system can pursue dynamic and static efficiency goals simultaneously.[8]

To summarise, the characteristics that were important in shaping issues at *project* level in innovative activity (i.e. uncertainty, non-specificity, lags and cost) have also proved central in analysing problems of R & D strategy and organisation design. Non-specificity, uncertainty, lags and cost are major considerations affecting R & D strategy and the incorporation of R & D within formal hierarchies, while non-specificity and uncertainty are particular features that may encourage adoption of an organic rather than a mechanistic mode.[9]

Thus, probing beneath the surface differences of conventional analysis of R & D behaviour, business and corporate strategy, and organisational form reveals interesting similarities and parallels in terms of the general relevance of the four common concepts. This encourages optimism as to the possibility of pursuing a policy of integrating the historically separate literatures, to a greater or lesser extent.

This line of possible development finds further justification in recent arguments by some organisational theorists that work in their field would be enriched and strengthened by adoption of an evolutionary perspective on the development of organisational forms (McKelvie and Aldrich, 1983). They argue that existing studies tend to suffer from over-simplistic assumptions, presuming that organisations are either all alike or unique. They argue that systematising analysis of organisational forms, and introducing notions such as variation, selection and competition into analysis, would create a more coherent analytical basis in this area of study.

Again, at a surface level, McKelvie and Aldrich's argument is symptomatic of the Balkanisation of analysis in this area, since they make no

reference to the extensive economic literature on evolutionary approaches to the study of organisation. However, at a deeper level, it suggests the possibility of commonalities and even convergence in terms of methods and frameworks between economics and organisational approaches, possibly complementing and building on the four common characteristics discussed throughout this chapter.

Concluding remarks

Non-specificity, uncertainty, delay and cost are important attributes affecting the economics of R & D activity. This chapter has attempted to show that it is possible to synthesise apparently disparate threads as far as issues in analysis of the R & D function, corporate strategy and structure are concerned. It is argued that analysis of function, strategy and structure should be rooted in the same core issues and problems.

We are also optimistic that it is possible to discern kinship relationships in approaches, such as evolutionary theory,[10] transaction cost economies[11] and organisational decision-making[12] in those subjects. If so, it offers interesting possibilities for cross-fertilisation between economic, business and organisational approaches.

Notes

1. This implicitly assumes that the specialised firm would not be able to appropriate a high level of gains through market transactions, e.g. joint venture and licensing. The reasonableness of this assumption may depend on the transaction costs associated with these alternatives in particular industrial environments.
2. The demand-pull approach is most commonly associated with Schmookler (1966). Mowery and Rosenberg (1979) have, however, questioned the legitimacy of many of the demand- or market-pull approaches to empirical interpretation.
3. However, the market still has an important role as an *ex post* selection device. See Nelson and Winter (1977) and Dosi (1982).
4. Freeman (1982, p. 167) points out that failure of divisional management to look far enough ahead was one consequence of the short-time perspective in profit-controlled divisions. Hayes and Abernathy (1980) identify the short-term orientation of divisional profit centres as a contributory factor towards what they perceive as the neglect of technological change in the US economy.
5. This was termed the Sappho project and is discussed in more detail in Freeman (1982, Chapter 5).
6. Since Burns and Stalker's seminal work, a great deal of theoretical and practical analysis has been undertaken regarding organic system design. Child (1984) provides an excellent coverage of both these aspects.
7. This does not mean that R & D lags and cost are not important in organic systems as they are in mechanistic. They will of course still be of relevance.

However, what the organic system does appear to offer is a particularly appropriate system for dealing with the other two characteristics of non-specificity and uncertainty.

8. See, in particular, Nelson and Winter (1982).
9. Williamson (1985) presents a recent statement of his development of transaction cost economics. Kay (1982, 1984) utilises transaction cost economics to analyse problems in corporate strategy and structure. For an analysis of problems associated with Williamson's existing framework, see Kay (1987).
10. For a survey of recent work in this field, see March and Shapira (1982).

References

Burns, T. and Stalker, G. M. (1968), *The Management of Innovation*, London, Tamstock.

Chandler, A. (1962), *Strategy and Structure*, Cambridge, Mass., MIT Press.

Child, J. (1984), *Organisation: A Guide to Problems & Practice*, 2nd edn., London, Harper & Row.

Dosi, G. (1982), 'Technological paradigms and technological trajectories', *Research Policy*, vol. 11, pp. 147–62.

Freeman, C. (1982), *The Economics of Industrial Innovation*, 2nd edn., London, Frances Pinter.

Hayes, R. H. and Abernathy, W. J. (1980), 'Managing our way to economic decline', in Steiner, G. A. *et al.* (eds), *Management Policy and Strategy*, 2nd edn., New York, Macmillan.

Kay, N. M. (1979), *The Innovating Firm, A Behavioural Theory of Corporate R & D*, London, Macmillan.

—— (1982), *The Evolving Firm: Strategy and Structure in Industrial Organisation*, London, Macmillan.

—— (1984), *The Emergent Firm: Knowledge, Ignorance and Surprise in Economic Organisations*, London, Macmillan.

—— (1987), 'Markets and false hierarchies: some problems in transaction cost economics, European University Institute Working Paper.

Knight, F. M. (1921), *Risk, Uncertainty and Profit*, Boston, Houghton Mifflin.

McKelvey, B. and Aldrich, H. (1983), 'Populations, natural selection and applied organisational science', *Administrative Science Quarterly*, vol. 28, pp. 101–28.

Mansfield, E. (1968), *The Economics of Technological Change*, London, Longman.

—— (1985), 'How rapidly does new technology leak out', *Journal of Industrial Economics*, vol. 34, pp. 217–23.

March, J. G. and Shapira, Z. (1982), 'Behavioural decision theory and organisational decision theory', in G. R. Ungson and D. R. Braunstein (eds), *Decision-Making: An Interdisciplinary Perspective*, Boston, Kent Publishing.

Mowery, D. and Rosenberg, N. (1979), 'The influence of market demand upon innovation: a critical review of some recent empirical studies', *Research Policy*, vol. 8, pp. 102–53.

Nelson, R. R. (1959), 'The simple economics of basic scientific research', *Journal of Political Economy*, vol. 67, pp. 297–306.

—— (1984), *High Technology Policies: A Five Nation Comparison* American Enterprise Institute.

14 Technological opportunities and industrial organisation

Rod Coombs

Department of Management Sciences, University of Manchester Institute of Science and Technology, Manchester

On the significance of technological regimes

During the 1960s and 1970s research on the details of the technical-change process tended to focus on particular innovations. One of the central questions studied was the relative importance of demand-pull versus technology-push as determining why particular innovations or clusters of them came about when they did. However, the idea of demand-pull was then much better worked out than the idea of technology-push. Perhaps partly because of this, most of the empirical studies were able to document the importance of the former but not the latter. The perception by scholars of the importance of demand-pull, and the ambiguity of the influence of technology-push, had a noticeable influence on the policy discussions, which tended to stress the possible roles of government in influencing the character of demand, and to downplay direct government R & D support.

The last decade has seen a significant increase in understanding, or at least a change in perception, regarding the nature and significance of 'technological opportunities'. The change has been associated with a change in the unit of observation from particular innovations to the development of technologies more broadly. Put another way, innovations came to be seen as connected, and the attention focused on the connecting structure, rather than on particular innovations in that structure. Nelson and Winter (1977, 1982) have used the terms 'technological regimes' and 'technological trajectories' to refer to the intellectual structure guiding technical change in a field at any time, and to the associated internal dynamics through which technology unfolds. Dosi (1982) has employed 'technology paradigm' to refer to the intellectual structure associated with a given technological regime. Sahal (1981) and Saviotti and Metcalfe (1984) have developed similar ideas. The underlying notion regarding all of these intellectual developments is that innovations should not be considered in isolation, but rather must be understood in terms of an evolving technological structure.

There has been a parallel shifting in thinking about national policies, from the stimulus of (more) individual innovations, to the fostering of advance of broad technologies, like information technology, biotechnology, etc. With this shift of focus has come more careful analysis of the appropriate roles of government and the finance of R & D aimed at advancing particular technologies.

These theoretical, empirical and policy developments have had influence on a wide variety of topics. One important one, and a central concern of this chapter, is the analysis of connections between technical change and industrial organization. During the 1950s, 1960s and 1970s, there was considerable research on this topic. However, almost all of it presumed that the causal flow was from industrial organization to technical advance. The notion of the technological regime which is common to a number of firms in the same industry calls attention to the possibility that causality may flow at least as much from the nature of technology and the character of technical change to industrial organisation, as from industrial organization to technical change.

The associated concepts of technological regimes, paradigms and trajectories have been explored in an earlier chapter. Here the focus is on the relationship between these notions and the connections between market structure and technological change.

Market structure and innovation: the two-way causality

The Schumpeterian tradition

The literature on the relationships between market structure, firm size and innovation is voluminous but has not been able to come to a firm conclusion. Both on theoretical grounds and on the grounds of casual observation it is easy to propose both advantages and disadvantages in innovation for monopolised and competitive market structures, and for large and small firms. In each case it is the unequal distribution of the incentive to innovate and the capacity to innovate which causes the ambiguity of prediction. Nevertheless, the force of the Schumpeterian position on market power and size, coupled perhaps with the industrial economists preference for exploring relationships in which conduct is derived from structure, has led to considerable empirical work testing this view. The two relationships most commonly tested are that research intensity increases with market concentration, and that research intensity also increases with firm size. Kamien and Schwartz (1982) review a large number of studies on these relationships. In the case of market concentration they find that there is little consensus, but remark that there appear to be strong industry effects and that the degree of *technological opportunity* at industry level may be a major influence on research intensity (first shown by Scherer, 1967). In the case of firm size, they suggest some consensus around the view that research intensity increases with size up to a certain point and then falls again. But once more they add the caution that the picture is probably more fundamentally affected by industry-specific variables which have not been included in the analysis. More recent work on market structure by Angelmar (1985), and on firm size by Rothwell and Zegveld (1981) and

Kaplinsky (1983), has confirmed that the patterns vary considerably between industries. This no doubt explains why no clear pattern is visible above the industry level of aggregation.

This conclusion—that variations by industry in technological opportunity are the primary variables explaining variations in research intensity—undermines the view that industry structure variables are a primary determinant of patterns in innovative behaviour. Indeed, it suggests that a strong causality runs from technological opportunity, through innovative behaviour, to industrial structure itself, thus reversing the previous orthodoxy (Momigliano and Dosi, 1983). In fact this represents a more general statement of an observation first made by Phillips (1971) in a case study of the aircraft industry; namely that rapid technical change and difficult imitation conditions in an industry tend to generate powerful pressure towards concentration. Subsequently, a number of other studies of particular industries have added weight to this view: see, for example, Katz and Phillips (1982) on computers, Dosi (1984) and Malerba (1985) on semiconductors, and Altshuler *et al.* (1984) on automobiles.

Industry life cycles

Somewhat separate in its origins from the debate on market structure and innovation, there has been a significant tradition of theoretical and empirical studies which have argued that many of the features of an industrial sector, including concentration, are related dynamically to the evolution of the dominant technology of the industry, in some form of product or industry life cycle. Variations in productivity growth rates between industries were explained in part by reference to the 'age' of the industry and its technical base by Salter (1966). Varying rates and interactions between radical product and process innovation were placed at the centre of the life cycle by Abernathy and Utterback (1975).

The detailed work of Walsh (1984) on the chemical industry and its subsectors from the middle of the nineteenth century to the 1970s has shown the subtle shifts in the nature of technological change in an industry which accompany its evolution. The early stages of the dyestuffs, plastics and pharmaceuticals industries were all characterised by bursts of discontinuous scientific and technical discovery, which were partly serendipitous, and partly resulted from broadly targeted scientific research activity. These were the Schumpeterian 'impulses' which led to reallocation of resources and structural change. As the markets for these products began to develop the character of technical change shifted to improvements, branchings and extensions of initial technical paradigms, and towards process change. Furthermore, the fine structure of technical change begins to reveal a closer 'tracking' of shifts in patterns of demand as the industries mature. This pattern of a technical *opportunity* being exploited through a group of technological *trajectories* is a very robust feature of industrial dynamics. It is a reflection of the path-dependence, or directionality, of technical change, which exercises such a powerful influence on the evolution of the industry itself. In later sections we explore further the evidence on the mechanisms which link these patterns of technical change to changes in industrial organisation.

We therefore see a contrast between two approaches to the relationships between technical change and industrial organisation. These two approaches emphasise different directions of causality between market structure and patterns of innovation. In part this arises from their differing starting points. In the traditional view, which emphasises the effects of market structure on innovation, the analysis tends to be fairly static, analyses incentives to innovate in different market structures, and is ultimately concerned with the normative issues of efficiency and welfare. Where models have been made more complex, to include such variables as retaliatory innovation and R & D spending, such as in the work of Dasgupta and Stiglitz (1981), the primary focus nevertheless remains on the issue of whether the effect of market power on innovative patterns is or is not socially desirable. The question of how that market structure has arisen is not addressed; indeed, in their basic model the number of firms in the industry and the R & D to sales ratio are determined simultaneously, rather than any causal relationship being suggested.

As we noted above, the main reason for questioning the traditional approach is the lack of fit with the empirical evidence. Stoneman (1983, p. 46) reviews several empirical studies which indicate that product market and technological characteristics seem to be operating alongside structure as determinants of R & D allocations and innovative behaviour, and that the specific combination of influences varies from industry to industry.

Those approaches which, by contrast, emphasise the causality from technical change to market structure, tend to take a more dynamic view of the issue. The *evolution* of industrial structure rather than its effects becomes the object of study, and hence the process of technical change, viewed as a sequence of related innovations, becomes a natural candidate for incorporation into the analysis. We now examine these issues more closely.

Stochastic growth models and industry structure

An interesting phenomenon in the field of industrial economics which is an important instrument for clarifying the relationship between technical change and market structure is the observation of serial correlations in the growth rates of firms. This has been observed in general studies of firm growth (Singh and Whittington, 1968) and has been given a more specific technical component by Mansfield (1968), who argues that successful innovators enjoy higher than average growth rates.

Nelson and Winter (1982) explore the causes of serial correlation by setting up and analysing a number of dynamic models. In all of their models, R & D aimed at innovation is distinguished from R & D aimed at imitation. Some firms spend on both, some only on imitation. The former are called innovators, the latter imitators. Both kinds of firms have target

R & D to sales ratios. As they grow or decline, their R & D spending does too. This is important because in these models the probability that an innovator will achieve an innovation is proportional to its R & D spending. A successful innovation gives a firm a production cost advantage over its competitors. The expected time before an imitator is able to latch on to the innovation is proportional to its imitative R & D spending. Thus in these models large firms are more effective innovators and imitators than small ones. In turn, the size of a firm at any time is related to its past innovative and imitative successes, since profitable firms are tempted to expand, although they may be dissuaded by concerns about spoiling the market, and unprofitable ones are forced to contract. The production cost advantages conferred by innovators can be set in the model to reflect either a regime of incremental innovation on the basis of the industry's existing technology, or a regime of radical, science-based technical change resulting in high rates of 'latent productivity growth'. Firms who have achieved size advantages can be modelled as behaving aggressively or with restraint in their investment decisions. Equilibrium is not a feature of the model.

The model takes the form of a computer simulation in which the initial conditions, including the number of firms, is set; and then the simulation is run for a number of decision periods. Amongst the results the following are particularly noteworthy.

1. In the simulations based on sixteen firms the level of concentration always increases over time.
2. The rate of increase of concentration and the final level achieved are modified by the rate of technical change and its regime, by the ease or difficulty of imitation, and by the degree of investment restraint of large firms.
3. Where the rate of technical change is fast, and imitation difficult, the tendencies towards concentration are most powerful, thus confirming the observation of Phillips (1971) in a more general sense.
4. Where firms in concentrated industries pursue aggressive growth policies, imitators tend gradually to achieve ascendancy over innovators. Where restraint is shown, innovators and imitators survive with innovators predominant.

Space precludes a fuller discussion of the detailed features of this model, but the results which are most salient to the issue of causality in the relationship between market structure and innovation can now be summarised.

Firstly, the model does demonstrate some traditional (and some more surprising) features of the causality which runs from market structure to innovation. Large firms in industries where small firms are also present do have innovative advantages which result in greater than average growth. Concentrated industries achieve a given rate of technical change for less total R & D expenditure, but give rise to a higher price level. Concentrated industries can shelter innovators if investment restraint is shown, or they can drive innovators out of business to the benefit of imitators if less restraint is shown, thus halting technical advance.

Secondly, however, and more striking, the forces of concentration are very strongly dependent on the rate and character of technical change. This confirms the original empirical finding of Scherer (1967) that differences in technological opportunity between industries are a more important factor than structure in explaining differences in innovative activity. Particularly interesting is the finding of a stronger result for the simulation of radical as compared to incremental technical change regimes. This is strong support for the Schumpeterian insight that the dynamics of industrial growth create powerful incentives for innovators and for imitators. Furthermore, it suggests a picture of a multiple set of contingencies affecting the two-way nature of the relationship between innovation and market structure.

Complicating factors

There are a number of complicating factors which now need to be considered, in order to move beyond the simplifying assumptions made in the discussion so far. The first of these, which is not incorporated into the Nelson and Winter model, is the question of entry. They acknowledge its importance (1982, p. 328) and argue that their results would not be fundamentally affected by entry of small imitative firms, but note that entry by large and technically progressive firms would change the picture considerably. In an extension of the basic model, Winter (1984) introduces a treatment of entry. He finds that in the 'cumulative' or incremental regime of technical change established firms can defend against entry better than they can in a more radical, 'science-based' technical regime. A more explicit treatment of entry is that of Gort and Klepper (1982) who offer a five-stage model of the patterns of entry during the evolution of a market as follows.

1. rapid growth of entry following the first innovation; conditioned by ease of entry and number of potential entrants;
2. increase in the number of producers;
3. entrants and exits cancelling out with net entry zero;
4. negative entry, or shake-out;
5. zero entry as the industry becomes mature.

This model has features in common with a number of other discussions of the industry life cycle. The essential principle is to contrast the nature of technical possibilities at the beginning and end of the process. At the beginning the technology is seen as volatile. External technologies possessed by potential entrants may facilitate entry, and disable existing producers. At the end of the process; technology-led entry is more difficult because the existing technology is mature, and the accumulated experience of the existing producers adds to the other mechanisms for deterring entry. This latter point is consistent with the analysis of Winter (1984). Gort and Klepper's work supports the view that the dynamics of innovation will

promote concentration over time, as in the Nelson and Winter model, though it starts from different assumptions. Furthermore it has significant empirical support. However, it perhaps oversimplifies the phases of evolution of the market by implicitly suggesting that the technology base of the industry will always be sufficiently well defined as to allow existing producers to maintain sovereignty over it into the mature phase of the industry. This may not be the case where there are multiple technologies embodied in the product.

If we acknowledge that complex products comprise a *range* of component technologies, and that radical change in one of them may alter the relative entry potential of outsiders, then the neat sequence of stages could be upset, or even periodically reversed. In other words, the contingent factors which affect the evolution of concentration, both in this model and in the Nelson and Winter model, are subject to variation not only *between industries*, but also *within* industries over time. Thus the evolution of market structure may not follow one clear path; but rather a zigzag of paths, reflecting variations in the conditioning factors of rate of technical change, ease of imitation, range of technologies employed in the industry, etc. We shall return to this point later.

A second complicating factor in developing the insights of the Nelson and Winter model and the Gort and Klepper model is the issue of economies of scale, economies of scope, and the multi-product firm. The latter point is particularly important since Nelson and Winter explicitly restrict their model (for good reasons) to single-product situations. Stoneman (1983) summarises the arguments on this point as follows. There is no reason to expect technical change to push the minimum efficient scale of plants in one direction or the other. However, the arguments of Williamson (1975) on the organisational benefits of the M-form company, and on its ability to utilise unexpected R & D gains, together with the advantages of such large firms in financing R & D, may have increased the minimum efficient scale of firms. If markets have not grown at the same rate then this process could increase concentration. However, since the flexibility of M-form firms is not infinite (and indeed the unrelatedness of diversification is correlated with diminishing performance) we can assume that such firms do not completely obliterate the significance of industrial boundaries, nor do they remove the upper bounds on concentration.

These issues have also been examined in a rather more general framework by Kay (1986). He argues that markets for capital, and for cooperative agreements concerning innovation, both between firms and within large diversified firms, play an important role in facilitating innovations in circumstances where product market structure would theoretically reduce incentives. Thus, for example, the availability of capital, and innovative opportunities, for deployment within diversified M-form organisations should reduce the effect of the apparent constraint on innovativeness implied by the structures of the product markets in which the organisation operates. Similarly, licensing agreements, joint ventures and other forms of agreement should in principle allow innovative opportunities to be realised in spite of market structure constraints. The caveat on both these mechanisms is that the markets in capital and in agreements have to function with sufficient efficiency to allow this flexibility of innovation to be realised. Clearly, this flexibility is not complete; but Kay argues convincingly that cooperation mechanisms and capital markets are able significantly to free the direction of innovative activity from the constraints which product market structure might otherwise impose.

Summarising this discussion of economies of scale and scope, we can say the following. The traditional analysis of the relationship between industrial organisation and technological change, in the context of the well-defined, single-product industry, hinges on the view that the capacity to innovate and the incentives to innovate are dissociated. The capacity lies with the larger firms and those with market power. The balance of incentive lies with the smaller firms. We have seen first that this dissociation is not black and white, and, secondly, that institutional innovations such as the multi-product firm can significantly *recombine* the capacity and the incentive to innovate. Thus the issue of industrial organisation becomes much more complex than in the traditional view.

The effect of this discussion of complicating factors such as entry, multi-product firms and collaboration, on our diagnosis of the innovation market structure relationship, is paradoxical. Whilst it strengthens the intrinsic importance of causalities flowing from technical opportunities to market structure, it highlights the fact that the sources of these technical opportunities, and the mechanisms through which they are applied, are complex and have multiple relationships to any industry being analysed. Clearly, then, the issue can only be resolved by recognising the need to define time periods, levels of aggregation, product characteristics, and bundles of technical characteristics, in such a way that the problem is constrained and made more manageable. For example, if we were to insist on conducting the analysis on an object defined as the transport industry, the problems would be immense, and any results might not be very meaningful. If we, however, restrict the problem to railways, or road haulage, or private cars, the issue becomes more tractable.

Technological opportunities and technological trajectories

The preceding discussion has identified the fact that the links between innovation and market structure can best be approached by seeing the problem dynamically, and by seeing innovations not as discrete, but as sequences of related innovations which intersect from time to time in complex ways. Technological opportunities come into existence, have a fruitful phase, and subsequently diminish in fruitfulness. These strongly directional processes are what have been variously referred to as technological trajectories, paradigms, etc.

If innovative opportunities do indeed arise in a structured form, in part as a result of the evolution of technological trajectories, then the foregoing observations suggest that product market structures will respond adaptively to them, through the behaviour of collaborating groups both within and between firms. Market structures will, however, exercise a significant influence on the rate at which trajectories are exploited, and new ones identified.

But, as was briefly mentioned above, a significant moderating influence should be taken account of at this point. Whilst technological opportunities and trajectories occur in the domain of *technology*, innovations occur in the domain of *products*. Products vary greatly in terms of their complexity; the range of technologies which they incorporate; and in terms of their sequential position with respect to other production and consumption activities. These variations condition the way in which related technological opportunities are realised in different types of product. Thus it is not possible to 'read off' a set of industry structure characteristics and dynamics directly from a diagnosis of technological opportunities. Proper attention must be paid to the range of interacting technologies relevant to the product or service, and to the position of the product in the input/output structure. This point can be illustrated through use of Pavitt's (1984) taxonomy of sectoral patterns of innovation. He identifies distinctive characteristics of the innovation process in 'supplier-dominated' sectors; 'production-intensive' sectors (sub-divided into 'scale-intensive' and 'specialised equipment suppliers'); and in 'science-based' sectors. The extent to which technical change is product- or process-centred, internally or externally generated, radical or incremental, varies between these types of industrial sector. It seems likely therefore that the patterns in the relationship between technological opportunities and market structures will vary between these types of industry. This represents a step in the direction of defining levels of aggregation and product characteristics which was earlier noted as a precondition for further clarifying the relationships between innovation and market structure.

It is also important, however, to consider the time period chosen for analysis of change in industrial structure. In the short to medium term, industries within Pavitt's categories may exhibit relatively stable patterns of causality. In the longer run, however, Pavitt's taxonomy from supplier-dominated through to science-based is one in which age of industry decreases and research intensity increases. Pavitt's categories of sector therefore have a dynamic relationship to technological opportunities as well as the cross-sectional relationship to which he draws attention.

In the next section we briefly consider technological trajectories a little further. One objective of the discussion is to identify those aspects of trajectories which are intimately connected to the experience and skills of particular firms, and contrast them with those aspects of trajectories which exist more independently, and which are available to groups of firms in industries. In making this analysis we are therefore assuming a frame of reference within theory of the firm which is essentially behavioural, drawing on the traditions of Penrose (1959) and Cyert and March (1963).

In so doing we are adopting a framework which is sympathetic to the microeconomic foundations of the evolutionary model of industrial dynamics discussed in this chapter so far. Earlier chapters in this book have discussed technological trajectories in more depth.

Firm-specific resources and technological opportunities in innovation

Institutionalised R & D departments have as their objective the creation of technology, but not of all possible technologies. Their possibilities are limited first and foremost by the objective characteristics of the knowledge bases within which they are working. A useful way of presenting this is in terms of the parameters which describe any particular system.

Most technical systems can be defined in terms of some performance characteristics which embody the function of the device and some physical specifications which determine the level of performance achieved. A simple example is the case of microelectronic circuits. One important performance parameter is the speed of operation of the circuits, which is of significance to users. This performance variable is a function of a specification, namely the degree of miniaturisation of the circuit. (The smaller the distance the electrons have to travel between parts of the device, the quicker the circuit can operate.) There is obviously a physical limit to this reduction in size. In this case it is determined by the laws of quantum mechanics which do not allow the device to function in a stable manner if certain elements are placed closer than a given distance. This limit is independent of any technical difficulties which may arise in achieving it. Other examples abound; the relationship between drag coefficient and fuel efficiency in motor vehicles is another one.

If the physical limit to any physical specification is still distant, and there remains considerable scope to operate on that specification to increase performance, then there exist substantial '*intensive*' technological opportunities in that system. By contrast, '*extensive*' technological opportunities will exist if the system in question has numerous possible functional applications in a variety of products or processes, where its performance characteristics will be of value. This latter is clearly the case with microelectronics. This approach to specifying technologies has been developed recently for the purpose of constructing indicators of technical change (Saviotti and Metcalfe, 1984). What is significant for the present discussion is that the technological frontier has some definite structure, and that technical reasons exist for expecting some strategies for altering that structure to be easier than others. This is a way of capturing the directionality of technical change.

However, the relative difficulty of different strategies for altering the technical frontier has to be expressed in terms of some unit; and the natural one is consumption of resources. It would seem appropriate to express some relationship between the cost of achieving some unit of technical change and the benefit which falls to the firm which makes the change.

Following Stoneman (1983), we can use the simple identity:

$$G = p{\cdot}x - c{\cdot}x - E$$

where G is profit
p is the price of the commodity incorporating the technical change
x is the level of output of the commodity
$x = sM/p$
s is the market share for the producer
M is the size of the total market
c is the cost of production of one unit
E is the expected cost of achieving the technical change

If a mark-up pricing policy is assumed then the expression becomes:

$$G = s{\cdot}M\,(l - k) - E$$

where $k = c/p$

Thus profit is positively related to size of market, share of market and mark-up, and is negatively related to the cost of achieving the technical change. This expression now allows some useful insights into the significance of different regimes for technological innovation.

Consider first the case of a radical innovation which opens a new market or serves an existing market with a product based on new technological principles. The technological opportunity, now measured in terms of the profit expectation, will be high where s and M are high; which may be the case if the performance or the cost of the new item offer substantial advantages with respect to previous products, and/or if entry barriers are low. But for the profit to be high s and M have to be high in relation to E. E will be low to the extent that the knowledge required to achieve the technical change is easily available to the firm involved. This will be a function of the R & D specialisation of the firm, and its collaborative possibilities.

Consider now the contrasting case of subsequent improvement innovations within the parameters of the technical system defined by the first innovation. These may increase the performance or reduce the price with respect to the initial state. The values of M and s for each incremental step will depend on the growth rate of the performance parameter. This will perhaps increase in the short run as a result of learning but will decrease in the long run as a result of encountering the intrinsic performance limit of the technical system, as described earlier. For these incremental innovations E may also diminish initially as a result of learning, but may increase ultimately as more complex methods are needed to modify specifications and increase performance. Thus the technological opportunities evolve in a manner which is related to the diffusion process initiated by the first innovation; an issue also analysed by Metcalfe (1981).

There are a number of overlapping concepts here. The first case discussed above—that of the radical innovation—closely resembles the concept of 'extensive' technological opportunities mentioned earlier, though the relationship is not exact. The second case—that of the succeeding incremental innovations—appears to be identical to the notion of 'intensive' technological opportunities and closely resembles the idea of a natural trajectory of technical change. Furthermore, the radical innovation/extensive opportunities scenario is essentially the same as the 'science-based and the cumulative regimes in the Nelson and Winter model are cussed above, and the incremental innovation intensive opportunities scenario corresponds to the 'cumulative' regime in that model.

This is an important conclusion which links the Nelson and Winter model with the industry life-cycle tradition. It is likely that the science-based and the cumulative regimes in the Nelson and Winter model are frequently linked together sequentially, given the appropriate choice of boundary conditions of time and product characteristics. An initial science-based regime would give a powerful 'internal' impetus to the process of concentration. A gradual shift to a cumulative regime can allow a relative attenuation of the causal mechanisms running from innovation to market structure, and an emergence of some of the reverse mechanisms. The Penrosian firm-specific character of much of the technical change in the cumulative regime can act as a defence against entry, as Winter (1984) suggests. However, this will be balanced by the enhanced danger of entry by science-based firms, especially if markets for capital and cooperation lower entry barriers. Hence a particular product function could become the site of a conflict between a decaying cumulative technical regime and an emerging new science-based regime. In the light of the increasing role of multi-product firms, and the strategic use of R & D, it is now possible for this conflict between technologies to be conducted not only through competition between firms but also through strategic choices made within firms.

These topics of firm-specific resources, theory of the firm, and technological trajectories are dealt with in more detail elsewhere in this book, and will not be developed further here. The purpose of this brief discussion has been to illustrate the argument that it is through this route that further progress can be made in exploring the structure and directionality of technical change. Such an exercise might then develop the insights into the links between directional technical change and market structure which emerged prominently from the models discussed in the earlier part of this chapter.

Conclusion

If it seems that the diagnosis in this chapter reduces our power to make general statements and to deduce public policy, then this is only partly true. The power to make certain kinds of general economic statements—such as 'oligopoly might be the optimum industrial structure for technical change'–is indeed reduced, since the theoretical basis for such statements has been undermined by the preceding arguments. But the power to make general statements about technological trajectories is reinforced, since their significance has been underlined by the arguments. This returns us to the point made in the first section, namely that policy-making as well as analysis has tended to swing back towards technology type as the focus of discussion and intervention in recent years. Despite the difficulties with making policy in this way, it seems to be the least unsafe method available.

It is worth recording another implication of the foregoing analysis which links to the macroeconomic sphere, also mentioned in the first section. If sectoral rates of technical change are relatively independent, and if they condition industry dynamics to a significant extent, then this closely resembles the supply side conditions of Pasinetti's (1981) approach to macroeconomics, in which the levels of output and employment growth are primarily determined by the structural change process, and equilibrium is only possible as a fluke. However, this does not preclude some linking of sectoral rates of technical change, through a mechanism such as Freeman's 'New Technology Systems', thus permitting long-run fluctuations in structural change, and hence in economic growth. Such connections between the industrial and macroeconomic dynamics of technological trajectories are also discussed elsewhere in this book.

References

Abernathy, W.J. and Utterback, J.M. (1975), 'A dynamic model of process and product innovation', *Omega*, vol. 3, no. 6.

Altschuler, A. *et al.* (1984), *The Future of the Automobile*, Cambridge, Mass., MIT Press.

Angelmar, R. (1985), 'Market structure and research intensity in high-technological-opportunity industries', *Journal of Industrial Economics*, vol. 36.

Coombs, R.W., Saviotti, P.P. and Walsh, V. (1987), *Economics and Technological Change*, London, Macmillan.

Cyert, R.M. and March, J.G. (1963), *A Behavioural Theory of the Firm*, London, Prentice-Hall.

Dasgupta, P. and Stiglitz, J. (1981), 'Entry, innovation, exit: towards a theory of oligopolistic industrial structure', *European Economic Review*, vol. 15.

Dosi, G. (1982), 'Technological paradigms and technological trajectories', *Research Policy*, vol. 11.

—— (1984), *Technical Change and Industrial Transformation*, London, Macmillan.

Freeman, C. (1982), *The Economics of Industrial Innovation*, London, Frances Pinter.

Freeman, C., Clark, J. and Soete, L. (1982), *Unemployment and Technical Innovation*, London, Frances Pinter.

Gort, M. and Klepper, S. (1982), 'Time Paths in the Diffusion of Product Innovations', *Economic Journal*, 92.

Hay, D. and Morris, D. (1979), *Industrial Economics, Theory and Evidence*, Oxford, Oxford University Press.

Kamien, M.I. and Schwartz, N.L. (1982), *Market Structure and Innovation*, Cambridge, Cambridge University Press.

Kaplinsky, R. (1983), 'Firm size and technical change in a dynamic context', *Journal of Industrial Economics*, vol. 32.

Katz, B. and Phillips, A. (1982), 'Innovation, technological change and the emergence of the computer industry', in H. Giersch (ed.) *Emerging Technology*, Tübingen, J.C.B. Mohr.

Kay, N.M. (1986), 'Industrial structure, rivalry and innovation, theory and evidence', Herriot-Watt University Working paper.

Malerba, F. (1985), *The Semi-Conductor Business: The Economics of Rapid Growth and Decline*, Madison, University of Wisconsin Press.

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

Metcalfe, J.S. (1981), 'Impulse and diffusion in the study of technical change', *Futures*, vol. 13.

Momigliano, F. and Dosi, G. (1983), *Tecnologia and Organizzazione Industriale Internazionale*, (1985), I Mulino, Bologna.

Nelson, R. (1985), 'Policies in support of high technology industries', Yale University, Institute for Social and Policy Studies, Working Paper 1011.

Nelson, R. and Winter, S. (1977), 'In search of a useful theory of innovation', *Research Policy*, vol. 6.

—— (1982), *An Evolutionary Theory of Economic Change*, Cambridge, Mass., Harvard University Press.

Pavitt, K. (1984), 'Patterns of technical change: towards a taxonomy and a theory', *Research Policy*, vol. 13.

Pasinetti, L. (1981), *Structural Change and Economic Growth*, Cambridge, Cambridge University Press.

Penrose, E. (1959), *The Theory of the Growth of the Firm*, New York, Wiley.

Phillips, A. (1971), Technology and Market Structure: A Study of the Aircraft Industry, Lexington D.C., Heath.

Rothwell, R. and Zegfeld, W. (1981), *Industrial Innovation and Public Policy*, London, Frances Pinter.

Sahal, D. (1981), 'Alternative conceptions of technology', *Research Policy*, vol. 10.

Salter, W. (1966), *Productivity and Technical Change*, Cambridge, Cambridge University Press.

Saviotti, P.P. and Metcalfe, J.S. (1984), 'A theoretical approach to the construction of technological output indicators', *Research Policy*, vol. 13.

Scherer, F.M. (1967), 'Market structure and the employment of scientists and engineers', *American Economic Review*, vol. 57.

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

Singh, A. and Whittington, G. (1968), 'Growth, profitability and valuation', DAE Occasional Paper No. 7, Cambridge, Cambridge University Press.

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

Walsh, V. (1984), 'Invention and innovation in the chemical industry: demand-pull or discovery-push?', *Research Policy*, vol. 13.

Williamson, O.E. (1975), *Markets and Hierarchies: Analysis and Anti-Trust Implications*, New York, Free Press.

Winter, S. (1984), 'Schumpeterian competition and alternative technological regimes', *Journal of Economic Behaviour and Organisation*, vol. 5.