Part VII Formal Modelling

Preface to Part VII

Gerald Silverberg
MERIT, Faculty of Economics, State University of Limburg, Maastricht

This volume is an attempt to articulate some of the dissatisfaction felt by at least some economists about the ability of current economic theory to deal with true economic change and development—in which technical change clearly plays a crucial role. And it certainly follows in a well-established, if somewhat heretical tradition in which the figure of Schumpeter stands out prominently. However, the heretical status of that tradition will remain its fate unless its positive insights can be translated into useful empirical hypotheses and a consistent and powerful theoretical framework truly appropriate to its task. In this respect the authors of this volume agree that it is time to go beyond the stage of lamentation and begin to lay the foundations of a new approach.

Taking its lead from the contributions in Part II on the need for a wider framework, the chapters that follow attempt to outline the formal, mathematical approaches which have already appeared scattered throughout the literature or are just now in the process of emerging, and demonstrate that they do indeed fall into a coherent pattern and represent a consistent alternative framework for doing economic analysis. Although the point of departure may originally have been dynamics, non-linearity, disequilibrium, stability analysis, selection, or the dialectic of chance and necessity, it is becoming clearer that these are all special aspects involved in the description of self-organisational and evolutionary systems. Our task is to identify the specific avenues of attack that will lead to a better understanding of the evolution of technologies, national economies and social relations situated at a deeper level than mere analogy. I hope the reader will be able to come away from this section with the feeling not only that these avenues exist, but that they lead to stimulating new insights into the economic process, many of which, though adumbrated in the past, begin to take on clearer contours when viewed in this perspective.

My own chapter attempts on the one hand to enumerate the different components of the overarching self-organisation framework while arguing for their essential interrelatedness, and on the other to make a prima-facie case for the relevance of this framework to economics. I then go on to examine in some detail a number of specific models which I classify under three headings: multi-equilibria and catastrophe-theoretic models, selection models and Schumpeterian dynamics, and models of the self-organisation of economic behaviour. There is a common thread underlying this sequence of models and it is extraordinary how a few basic mathematical structures can be adapted to such seemingly disparate applications when properly handled.

Metcalfe then focuses on the diffusion of innovations as an exemplary process at the core of technical change. After examining the equilibrium tradition in this field, he opts for an evolutionary approach, which leads him to a discussion of a number of selection models in some detail. These models progressively incorporate more detailed features of market competition and complement my discussion of the strengths and weaknesses of current selection models. Finally, he analyses what he calls the process of Marshallian diffusion, which allows the explicit derivation of substitution curves as a function of the characteristics of a new technology.

Arthur's chapter on competing technologies goes beyond this discussion of diffusion/selection by introducing a significant additional feature: a non-linear dependence of the relative rates of growth (or the probabilities of adoption) of the technologies on their present or even expected shares, whether due to learning, standards, increasing returns, etc. This apparently small change in the dynamics of the competitive process is significant for two reasons. First, it is a prime example of a collective phenomenon, in which the decision of the individuals is constrained by the collective in such a way that several possibly exclusive alternatives contend for dominance. Second, it underscores the crucial role of small historical events which can trigger the eventual choice between these alternatives. What is particularly impressive about Arthur's chapter is the invocation of a very general and rigorous analytical result on the asymptotic behaviour of non-linear stochastic processes. This result enables us to analyse a wide class of increasing returns phenomena on the basis of only a qualitative understanding of the relationships involved, and is a powerful complement to simulation results.

One implication of the evolutionary modelling presented in this section is the doubt it casts on the unambiguous nature of 'optimal' behaviour and strategic rationality. Lock-in to an inferior technology is also seen to be a real possibility, which may require action going beyond the capacity of individual agents acting without coordination to overcome.

Whereas Arthur and Metcalfe focus on the industry-level dynamics of technological evolution, Boyer's chapter is an attempt to integrate several different forms of technical change with distributional mechanisms to obtain a dynamic model of an entire economy. Taking his cue from the French '*régulation*' school to which he has been a principal contributor, the analysis focuses on a number of fundamental 'regimes of accumulation' representing feedback networks between productivity, wages, consumption, investment and employment. He establishes links between these basic components by drawing on Schumpeterian and Kaldor–Verdoornian insights concerning productivity growth, and a generalised Phillips curve with respect to real wages. Qualitative changes in the resulting growth patterns are related to different constellations of the underlying parameters, which Boyer associates with long-term structural changes in the economy.

Obviously, this can only represent a simplified attempt to reflect important features of technical change and investment feedbacks in an overarching model of economic development. It remains a task for the future to work out a framework for analysing the economy-wide repercussions of concrete innovations in historical time, whether they be 'purely' technological or social as well.

---

24 Modelling economic dynmaics and technical change: mathematical approaches to self-organisation and evolution

Gerald Silverberg
MERIT, Faculty of Economics, State University of Limburg, Maastricht

Introduction: self-organisation and economic change

In this chapter we shall discuss some relatively new approaches to mathematical modelling which may loosely be subsumed under the heading 'theory of self-organisation'. Although this modelling philosophy and most of its early applications were originally inspired by problems in the natural sciences, as we shall argue, its relevance to the social sciences and in particular to questions of economic development and structure is more than accidental.

The theory of self-organisation deals with complex dynamic systems open to their environments in terms of the exchange of matter, energy and information and composed of a number of interacting subsystems. Thus the 'behavioural environment' and the individual subsystems are conceived as undergoing a process of mutual coevolution which may admit a determinate joint outcome. Within certain domains, in particular, in the neighbourhood of a structural instability, these interactions can often be represented at an aggregate level by a small number of *order parameters* which summarise the net result of the complex of feedbacks constraining the behaviour of the subsystems. Many such systems have been shown both experimentally and theoretically to lead to the spontaneous emergence of coherent macroscopic structures (e.g. spatial, temporal, or in terms of other system attributes) from the seemingly uncoordinated behaviour of the component parts at the microscopic level. Moreover, self-organising systems can undergo a succession of such structural transformations in response to generalised changes in outside conditions coupled with internal fluctuations at the microscopic level. In some cases this can take on the character of an evolutionary progression. Good overviews of the field can be found in Ebeling and Feistel (1982), Haken (1983a), and Nicolis and Prigogine (1977). The relevance of the self-organisation concept to the social sciences has been discussed in Prigogine (1976) and Prigogine, Allen and Hermann (1977).

Before turning from this rather abstract description to details of specific methods and applications, one may rightly ask what makes the concept of self-organisation of interest to economic theory, and in particular to the incorporation of processes of technical change and economic development at the very centre of its reformulation. To begin with, the economic system in a biophysical sense is certainly open, dependent on inputs of energy and information to maintain the processes of circular flow traditionally analysed by economic theory. This point has been emphasised in particular by Georgescu-Roegen (1971, 1976) as well as by Boulding (1978, 1981), and even earlier by Lotka (1924) and Marshall (1890). This means that in some fundamental sense the laws applicable to the general process of biological evolution and ecological interaction will have their counterparts in the economic realm. This is not a question of superficial analogy, however, asserting a one-to-one correspondence between biological/physical phenomena and economic ones. Rather, it implicates similar causal patterns of, for example, competition, cooperation, and the generation of variety operating in the 'deep structure' of both systems.

Second, it has become more clearly realised only in the last few years that both the mathematical richness and the empirical realism of the study of dynamical systems increases immeasurably when the focus shifts to intrinsic *non-linearities*. In economics this insight goes back to Richard Goodwin (1951), who showed that self-sustaining business cycles were only possible in the context of non-linear models. In addition to self-sustaining cycles, non-linearity introduces the possibility of systems with multiple equilibria, bifurcation of solutions of various types, and deterministic chaos, i.e. systems which, although deterministic, demonstrate no long-term regularities of behaviour and are highly sensitive to the choice of initial conditions. As we shall see, these features play an essential role in the mathematical theory of self-organisation and evolution. Conversely, it can be shown that non-equilibrium open systems can only display evolution if they are in the non-linear region. Once such non-linearities are admitted into economic modelling many traditional equilibrium approaches are called into question, while some qualitative thinking which has eluded formalisation until now can be given mathematical expression. Thus the theory of self-organisation also addresses the fundamental question raised by Adam Smith in economics: how do coherent market solutions emerge from the uncoordinated pursuit of self-interest of individual agents? But it makes clear that there may be a variety of answers to this question, many of them possibly suboptimal, and dynamically non-trivial.

A further departure from orthodox modelling philosophy, but one which also marks a reopening of scientific thought to historicity and the unique role of events, acts and individuals, is the place of stochasticity and irreversibility in processes of self-organisation. In contrast to mainstream econometrics, for example, which attempts to uncover unique structural laws from under the veil of stochastic noise, which simply serves to obscure them, stochasticity—the deviation of components and subsystems from mean values—is dialectically intertwined with deterministic regularities possibly to drive the system along new branches structurally distinct from past regimes. In biology this is what Jacques Monod (1970) called the interaction of chance and necessity. In economics it is not unrelated to the observation that a limitation of the large econometric models is their breakdown in the face of structural change, which they are not able to anticipate. The dialectic of chance and necessity impinges on the fundamental problem of the emergence of novelty. As against equilibrium models which see the economic process as one of adjustment to given conditions, which then may change for exogenous reasons and be continuously and almost timelessly tracked by the system, the theory of self-organisation examines the conditions under which departures from prevailing behaviour can become self-amplifying and modify the very environment hitherto dictating that behaviour. Since these departures are strictly speaking unpredictable at the macroscopic level, the cost is in terms of the precise ability to predict when, if and exactly which novelties may exert a significant effect on the system. The gain is, first, the open admission that the social sciences are indeed historical, but, second, the possibility of making educated statements about the kinds of change that may take place, patterns of regularity when it does, the existence of competing scenarios, and the magnitude of efforts needed to trigger a choice. This, of course, is the very stuff of political economy and economic history, which may now be open to an appropriate form of analytical treatment and more resistant to the kind of ideological scientism (in the sense of Hayek) model-guided theorising has been accused of in the past.

Finally, research on self-organisation has focused attention on the critical role of *collective phenomena* and *cooperative effects* in many systems. These features are of special relevance to applications in the social sciences and to the question of the relationship between individual behaviour and 'rational' choice, on the one hand, and the socio-economic environment (climates of opinion, social norms and institutions, herd effects, etc.), on the other. Systems displaying these properties converge to one or the other aggregate state depending on the distribution of initial states of the subsystems and possible thresholds and triggering events constraining deviant components to align themselves with the rest of the system. In a sense the system is able to 'vote' itself into a more structured or differentiated pattern because of the strong non-linearities mediating distributions of behaviour. Modelling approaches that proceed from the concept of a representative agent and unique self-consistent behavioural equilibria completely miss this point and thus are only able to account for institutional patterns of behaviour, implicit forms of cooperation apparently at odds with myopic self-interest, and pathologies like speculative bubbles, in highly artificial and *ad hoc* ways, if they recognise them at all.

In the following sections of this chapter we will illustrate these principles using a number of models which have been analysed in the economic, social science and biological literature. We will not present a detailed background to the mathematical methods themselves but instead refer the reader to the appropriate sources in the literature. Roughly speaking, these models fall into two categories. The first demonstrate how a social system moves between a small number of qualitatively distinct dynamic states, either cyclically or in response to input variables. The second show how such systems may encounter critical switch points which progressively and irreversibly drive them down a branching tree of specific development paths. This second class of models exemplifies the transition from self-organisation to evolution.

Multiequilibria and catastrophe-theoretic models

Consider a (multidimensional) dynamic system with state space vector $\mathbf{x}$ parameterised by a vector $\mathbf{a}$:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{a}).$$

If $\mathbf{f}$ is a linear function of $\mathbf{x}$ then in general only a unique stationary state X_0 is possible. Its stability properties depend on the eigenvalues of the matrix $\mathbf{f}$: if their real parts are all negative (or at least one is positive) then $\mathbf{x}_0$ is asymptotically stable (unstable) (see Hirsch and Smale, 1974, for a thorough introduction to dynamic systems and the associated linear algebra). If, however, $\mathbf{f}$ is a *non-linear function of* $\mathbf{x}$ then more than one stationary state can exist (for the moment we will leave aside the question of the existence of other 'attractors', i.e. other subsets of the state space such as closed curves, tori, and so-called strange attractors of fractal dimension invariant to the dynamic and exerting an 'influence' on other trajectories of the system). Their stability properties can be determined by linearizing the non-linear function $\mathbf{f}$ in the neighbourhood of the stationary points and examining the corresponding eigenvalues as in the linear case. In general, both the number and the stability of the stationary points may change as a function of the parameter vector $\mathbf{a}$. The values of $\mathbf{a}$ at which such qualitative changes take place are referred to as bifurcation or catastrophe points. (In the mathematical literature one proceeds with greater generality by analysing the topological structure of the flow, i.e. the ensemble of all trajectories generated by $\mathbf{f}$, and determining for what values of $\mathbf{a}$ it changes. The concept of *structural stability* goes further and parameterises the system with respect to all possible small perturbations of the equations. The system is structurally stable if the topological structure of its flow is invariant with respect to all sufficiently small perturbations of a certain class. Otherwise it is structurally unstable.) For a certain kind of dynamical system, namely *gradient* systems, it is possible to classify completely the kinds of bifurcations that can take place locally for parameter spaces of dimension less than or equal to four (cf. Poston and Stewart, 1978; Saunders, 1980; Thom, 1975). Gradient systems are characterised by the fact that they are derivable from a scalar function $V(\mathbf{x},\mathbf{a})$ (analagous to potential energy in mechanics):

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{a}) = -\operatorname{grad} V(\mathbf{x},\mathbf{a}) = -(\partial V/\partial x_1, \partial V/\partial x_2 \ldots \partial V/\partial x_n).$$

The stationary points correspond to the extrema of V, with maxima being unstable and minima stable. One-dimensional systems are always gradient systems, but in higher dimensions gradient systems are a very restrictive special case. If we are only interested in the behaviour of the stationary states in response to changes in the parameters (thus regarding them as input or control variables) as an exercise in comparative statics, or if we can assume that the relaxation of the system to equilibrium occurs considerably more rapidly than the variations in the control parameters (the fast/slow or adiabatic approximation, of which more later), then catastrophe theory becomes applicable. Given the dimensions of the state and of the parameter space, the classification theorem says that if a qualitative change takes place, locally it must be topologically equivalent to one of a small number of 'canonical' polynomials relating the equilibrium values of the state space variables to the parameters. Topological equivalence means that the original parameters of the systems may have to be transformed via a possibly complicated but invertible and continuous mapping to the canonical variables. (This is a point that is often overlooked in naïve applications of catastrophe theory.)

Zeeman (1974) furnishes a non-trivial example of the kind of argument which can legitimately invoke the classification theorem and how it can be embedded in a more complete dynamic model. He considers the dynamics of a stock market under the influence of two kinds of agents (or one kind of agent with two differently weighted and possibly conflicting motives): fundamentalists, who orient themselves around some notion of a natural price (e.g. using price/earning ratios), and speculators, who react to the direction of change of market price levels. Depending on the proportion of money in the market of the two types of investor, Zeeman argues that the market will be characterised by either one or two possible equilibrium states. This is due to the fact that when speculative money dominates, speculators as a group will reinforce swings in price movements. This leads to a cusp structure for catastrophes in the parameter plane marking the transition from the region with a single stable equilibrium to one with two stable and one unstable equilibrium. As with many of the early catastrophe-theoretic models, Zeeman assumes that his economic parameters can be identified directly with the canonical cusp parameters. He then goes a step further by imposing a 'slow' cycling motion in the parameter space representing a market successively dominated by fundamentalists and speculators. The sudden catastrophic jumps correspond to the crisis typical of the switchover from a bull to a bear market. Characteristic features of catastrophe-theoretic models are, for example, hysteresis and sensitive dependence on intial conditions. Hysteresis means that when a trajectory in parameter space which crosses the catastrophe set and induces a jump in the system is reversed, a return jump is not induced at the same point but only later when another part of the catastrophe set is traversed. The system thus acquires a primitive path-dependent memory of its own past. Trajectories in parameter space may also split, i.e. two trajectories starting out very close together but running along either side of, say, the cusp point of the catastrophe set can lead to widely divergent system states over time. Thus certain initial configurations can be very crucial for the evolution of the system and can magnify the effect of random events occurring near such parameter values, enabling them to decide which divergent path the system will be committed to. Random events can also be decisive near 'overhanging cliffs' of the catastrophe surface, tripping the system over the edge so to speak and thus effecting a rapid switchover from one solution surface to another.

Another multi-equilibrium model of direct relevance to the macrodynamics of technical change was first presented by Mensch *et al.* (1980) and subsequently reformulated and extended in Haag, Weidlich and Mensch (1985). This model was inspired by the observation that the relationship between investment, employment and national product seems to have broken down in the early 1970s. A number of authors suggested that this had something to do with the composition of investment and not only its absolute magnitude. Thus investment in modernising, rationalising, or accelerated replacement of equipment (possibly due, for example, to investment in more energy-efficient machinery in the wake of the oil crisis) would have a different effect on employment than pure expansionary investment. This fact had not been explicitly taken into account by macroeconomic models. (In fact, most models assumed that replacement investment was a fixed percentage of the capital stock, something which can only be justified in the long run in a golden age, steady-state growth universe.) Mensch *et al.* hypothesised that for certain combinations of expansionary and rationalising investment the economy would be characterised by two short-run equilibria representing full and underemployment. Under this assumption the simplist realisation is a cusp catastrophe with the levels of expansionary and rationalising investment (suitably normalised) as input parameters and the level of activity of the economy as state variable (assumed in short-run equilibrium). As the ratio of rationalising R to expansionary E investment increases, the (approximately) linear relationship between activity X and E becomes two-sheeted over a certain range. An important implication of such a model is related to the phenomenon of hysteresis mentioned above. In the bi-equilibrium region, once the economy has switched from the high to the low activity sheet, expansionary investment will have to be raised to a much higher level to trigger a spontaneous return to high activity than was necessary to keep the economy on the upper sheet. This may be the key to explaining why employment programmes often generate disappointing and only temporary increases in employment unless they exceed certain threshold levels.

The 1985 reformulation of the model differs from the original version in that (a) the procedure for identifying and incorporating the relevant input variables has been enlarged, and (b) short-period dynamics are also taken into consideration. The basic idea of a possible bi-stability in the system is retained, however. This leads to the construction of a fourth-order polynomial potential function with two time-dependent parameters. The potential function is estimated by least squares after filtering out fluctuations with an averaging process. This yields a time series for the two input variables, as yet unidentified. The relevant economic inputs are identified by assuming that the input variables in the potential function are linear combinations of variables selected from a set of possible candidates (including lags). The authors performed a correlation analysis on combinations of total investment, expansionary and rationalising investment, an 'investment structure index', open positions in industry, working hours in industry, and the rate of price increase. The best fit to the empirically estimated input variables for West Germany was obtained for linear combinations of expansionary investment lagged one year and rationalising investment lagged three years. For a thorough discussion of problems of estimation in catastrophe-theoretic models (in contrast to standard econometrics, bi- or multi-modal error distributions have to be assumed) in the context of a model of inflation by Woodcock and Davis (1979), see Fischer and Jammernegg (1986).

One drawback of much of the modelling inspired by catastrophe theory is the distinction between state variables and parameters. For example, in the two versions of Mensch's bi-equilibrium model, the components of investment are regarded as exogenous variables. In a complete dynamic description, however, it is clear that a feedback of some form exists between economic activity and investment. This is acknowledged in Zeeman's model by the introduction of the slow, second level of dynamic interaction between all the variables. The assumption that the motion can be clearly divided between fast and low responses permits the state variable/parameter distinction to be imposed mathematically, however, at least as an approximation. This procedure, known in physics as the adiabatic approximation, is the basis for Haken's so-called slaving principle, which establishes a hierarchy of causation between order parameters and 'slaved variables' near an instability in complex interdependent systems (cf. Haken, 1983a and b). Needless to say, the bifurcations involved can be of a more general type than elementary catastrophes. The distinction between slow and fast variables is implicit in most comparative statics exercises in economics. It is often assumed that some set of variables is in equilibrium while another set of parameters can be freely varied to represent economic change. The dangers of this kind of implicit dynamic reduction have been pointed out by Gandolfo and Padoan (1984). If one makes no *a priori* assumptions about the adjustment speeds of the various dynamic interactions and actually estimates them against the data, it turns out that the equilibrium assumptions of, for example, capital market clearing models are revealed to be untenable.

Selection models and Schumpeterian dynamics

Inspired by the original contributions of Schumpeter (1919, 1947) and Alchian (1951), a rapidly growing number of authors have attempted to model formally economic competition and growth, technical choice and diffusion, and technologically induced fluctuations as an *evolutionary* process. A glance at the biological literature shows that evolution is characterised by (a) selection of superior types from a heterogeneous population (the almost meaningless tautological phrase 'survival of the fitness' is replaced by the observation that environmental pressure does in general compel selection (cf. Eigen, 1971)), and (b) being open-ended and driven by the continual creation of variety originating in a primarily stochastic mechanism.

The economic models to be discussed in the following share a very similar methodological point of view and employ variants of the same basic mathematical structure. This mathematical structure has been termed *replicator dynamics* (Schuster and Sigmund, 1983) and has been at the centre of research in such seemingly diverse fields as sociobiology, prebiotic, macromolecular evolution, population ecology, and recently game theory as well. The basic equation was first introduced by R.A. Fisher in his mathematical formulation of natural selection:

$$\dot{x}_i = \mathrm{A}x_i[\mathrm{E}_i - \langle \mathbf{E} \rangle], \quad i = 1,n,$$

where

$$\langle \mathrm{E} \rangle = \Sigma\, \mathbf{x}_i E_i.$$

x_i represents the proportion of species i in some population of interacting species and E_i its related 'reproductive fitness'. $\langle E \rangle$ is the reference average fitness level of the population. The frequency of a species grows differentially according to whether it is characterised by above- or below-average 'fitness', while average fitness itself varies in response to changes in species frequencies.

The case originally investigated by Fisher was for constant E_i's. He showed that the system monotonically converges to a pure population consisting of the species with highest fitness. In the last few years considerable attention has been devoted to systems with quadratic or cubic dependence of the E_i's on the frequency vector $\mathbf{x}$ (and thus incorporating more complex feedbacks between the species such as the interesting case of cyclic interaction). Recent results are surveyed in Hofbauer and Sigmund (1984), Sigmund (1986), and Ebeling and Feistel (1982). The level of interactive complexity can be taken a step further by introducing additional dynamic variables $\mathbf{y}$ and lagged values of some of them $z_i(t) = y_i(t-\Delta)$, $i = 1,q$:

$$E_i = E_i(x,y,z)$$
$$\dot{y}_i = f_i(x,y,z),\ i = 1,p.$$

This type of system is necessary to model the vintage structure of capital stocks in disequilibrium and with possibly fluctuating rates of best practice technical change.

The models developed along these lines differ in a number of important respects, however, which can be roughly summarised under the following headings.

Unit of selection: Whereas the gene has come to be recognised as the fundamental unit of selection in biology, it is still unclear at what level evolutionary selection and innovation operate in socio-economic systems. In terms of the Schumpeterian model of creative destruction, for example, it is not obvious whether the basic unit should be the firm, or the innovation or technology itself. In addition, one may attempt to model behavioural strategies, rules of thumb, etc., as subject to an evolutionary process. All of these approaches are represented in the literature. It remains to be seen to what extent they can be reconciled, or whether they introduce an implicit bias into a model.

Behavioural assumptions and the role of anticipation, planning and 'rationality': Almost without exception, workers in the field of social evolution acknowledge that human societies are characterised by an emergent property almost totally absent from the biological domain—the presence of conscious goal-seeking behaviour partly guided by mental models of the world which attempt to anticipate the future course of the individual's environment. An extreme position might regard this fact as irrelevant to the ultimate outcome of the evolutionary process and therefore would dispense altogether with a detailed treatment of the behavioural level (Alchian, 1951, at times argues in this vein, and it seems to be implicit in the work of Marchetti, 1983). However, even under this assumption it would not be without interest to investigate the behavioural level as experienced by the actors themselves, if only as a problem in social psychology. Moreover, even if the outcome remained the same, the fact that the search process is not wholly random but directed, i.e. *orthogenetic* (cf. Lotka, 1956, p. 379), implies that it may be advancing much more rapidly than blind biological evolution. However, it is generally accepted that the behavioural level plays an essential role in the socio-economic process. One need only point to the importance of *imitation* in human affairs, which implies that successful strategies can be transferred between living agents and do not necessarily have to drive the carriers of other strategies physically out of existence. (As far as I am aware, only bacteria practise an analogous direct transfer of genetic information without reproduction.) And this example makes clear that the prevailing conception in economics of behaviour as 'rational' is woefully inadequate as a description of human beings interacting in a social and historical setting.

Unfortunately, it must be admitted that the behavioural level has been relegated to a mostly *ad hoc* part in most of the economic models, with the exception of some of the work of Nelson and Winter on the selection of decision rules. Particularly in the models based on evolution in technology as opposed to firm space, behavioural assumptions about, for example, the distribution of investment are often based neither on genuine profitability calculations of some kind nor on the sort of concrete decision procedures which have in fact been uncovered in the survey literature. Indeed, one of the strengths of economics over sociobiology, for example, is that the investigator can actually ask agents what influenced their decisions and in many cases expect a reasonably revealing reply, whereas even today the biochemical connection between genes and manifested behaviour in animals is little more than a fruitful heuristic hypothesis.

Phenotype vs. genotype: Although I have argued that reasoning by analogy is a less promising approach than the search for structural isomorphism, it is illuminating to try to apply for a moment this fundamental biological distinction to economics. It is not enough to locate units of competition such as technologies, firms' pricing, investment, or R & D policies. It is also necessary to specify in detail the *economic mechanism* governing their competitive interaction, which may be located at a somewhat different level of the system (the phenotypic level, so to speak). Thus it is a somewhat surprising fact that many of the selection models purporting to be descriptions of market competition do not have any but a rudimentary economic means of translating the underlying diversity of techniques or behaviour into the competitively relevant variables such as prices, production levels, delivery delays, product quality, etc. The evolutionary process does not simply work straightforwardly in both directions between the genotype and the phenotype, as the example of sexual reproduction and dominance makes clear. Too few of the models presented until now devote enough attention to the intervening variables mediating the process of economic competition.

Explanatory power of evolutionary modelling: Quite aside from such questions as superior correlative fit or predictive ability, it is important to ask what one hopes to achieve with this approach that is not attainable within the prevailing equilibrium/optimisation paradigm or with the use of specific *ad hoc* models. Simply to use evolutionary modelling to reproduce the common currency of orthodox theory strikes one as too modest a programme to justify the theoretical detours involved, even if the evolutionary approach may claim to be in some sense more realistic or plausible. One answer may lie in simpler and more robust solutions to such outstanding economic problems as oligopolistic pricing, which has not yielded up its secrets in the form of a usable dynamic description despite having been subject to an impressive assault with the sophisticated methods of game theory over the years. And this although the number of possible solutions in reality appears to be quite restricted and certainly smaller than the number of theoretical solution concepts proposed thus far. Problems such as these fall under the general heading of the origin and stability of cooperative behaviour, a subject which only recently has begun to be the focus of analytical investigation along evolutionary lines (cf. Axelrod, 1984).

Another goal to which many of the evolutionary models about to be discussed have addressed themselves is to uncover long-term patterns of technical change and economic development. One such result is the familiar logistic substitution curve of technological diffusion, but the question remains of how to embed it in a more general theory of economic dynamics. Another is the possible existence of long-term macroeconomic fluctuations and patterns of structural change, and the still highly contested long-wave hypothesis. Still another is the origin and persistence of patterns of unequal development in the world economy. One of the most promising questions, however, which has hardly been brought out in the literature, is the relationship between the process of technical change and the disequilibrium structures it engenders on the one hand, and short-period instabilities and the problem of effective demand on the other. This is one of the missing links (between Schumpeter and Keynes, so to speak) in economic theory which the equilibrium paradigm is singularly unsuited to deal with. On this more later.

To illustrate the basic form of the evolutionary argument let us start with the model in Nelson (1968) and Nelson and Winter (1982, pp. 235–40). While it is conceived to deal with underdeveloped economies characterised by a modern and a traditional industrial sector, in principle it is equally applicable to any closed economy which admits such a bipartite representation of capital-embodied technology. The technologies, assumed to be linear production functions, differ only with respect to their labour productivities. Each sector is wedded to its technology and reinvests its profits in capacity expansion (the rate of capacity utilisation is always one, output sells for a constant common price, and the wage rate is determined by a static labour supply curve), which results in differential growth rates of the two sectors. The more productive technique gradually replaces the less productive one in an approximately logistic fashion, depending on the form of the wage function. The rate of replacement is proportional to the difference in labour productivity.

A generalisation of this approach is presented in Silverberg (1984) based on Goodwin's growth cycle model (Goodwin, 1967). The argument consists of two parts: a hypothetical Gedankenexperiment and a descriptive dynamic analysis. The first part proceeds from the single linear technology Goodwin model and asks how the choice of technique problem can be answered dynamically in a Schumpeterian sense. That is, it asks under what circumstances an entrepreneur investing in a new technology of general linear form will eventually realise differential profits and establish himself in the economy. Using a method of analysis first applied by Allen (1975, 1976) to biological evolution in ecological systems, it is possible to derive an unambiguous selection criterion which is independent of factor prices prevailing at any given time. The well-known stylised facts of economic growth—approximate constancy of the capital/output ratio and the progressive increase of labour productivity—are shown to be necessary consequences of the assumptions underlying Schumpeterian competition, even if innovating entrepreneurs search for new technologies in a completely arbitrary manner. Thus the directed search for primarily labour-saving technologies is a logical and self-consistent consequence of this result.

The second stage of the analysis goes on to make the heroic assumption that the further evolution of the economy can actually be described by strict reinvestment of the respective retained earnings of the technologies. There is obviously a certain irrationality to an entrepreneur's continuing to invest in a demonstrably inferior technology over a considerable period of time, but it is not an entirely unknown phenomenon in business history. Under this assumption it is possible to derive an analytical expression for the substitution process of logistic type whose speed is proportional to the difference between the values of the technical choice function of the two techniques. Furthermore, the time paths of such macroeconomic variables as the rates of unemployment and average profits and the rates of growth of real wages and product result from a superposition of long- and short-period fluctuations.

A number of models have continued in a similar vein and examined the case of an arbitrary number of competing techniques under similar behavioural assumptions. Closely related are Gibbons and Metcalfe (1988) and Nelson and Winter (1982, pp. 240–5). In essence these are industry-level analyses because wages and other factor prices are taken as exogenous. The strict technology reinvestment assumption is retained. One may then demonstrate that for given factor prices and constant technology sets a best technology exists and the industry progressively converges to it. The rate of convergence of average industry unit costs to best practice is proportional to the variance of unit costs of the technologies present in the industry (using market shares as weights). If factor prices change exogenously, the best technique (in the sense of lowest unit costs) changes accordingly, and, depending on whether sufficient variety remains in the industry, the system now converges to that technique. This result is known in population genetics as Fisher's 'fundamental theorem of natural selection' (Fisher, 1930). For a good discussion of the underlying mathematics, see Ewens (1979), Hofbauer and Sigmund (1984), and Losert and Akin (1983).

The restrictive assumption of strict reinvestment in each technology can be relaxed by allowing investment to 'climb the ladder' of available technologies gradually over time, preferably in the direction of the techniques which are being selected for anyway. Examples of this sort of model are Soete and Turner (1984) and Iwai (1984a). The underlying assumption revolves around the idea that information about best practice diffuses only slowly, so that firms have to work their way up the production possibility set in a somewhat random fashion. This sort of process can be generated by positing that the rate at which profits earned on technique i are shifted to investment in technique j is proportional to the percentual difference in current rates of profit of the two techniques times the share of technique j in industry capacity (an imitation effect). This relationship is definitely situated in technology rather than firm space and is not clearly connected to any firm-level behavioural decision rule. (Thus two firms using identical technologies are completely free to pursue totally different investment strategies, a fact which is aggregated out of this kind of model.) The net effect of this assumption is simply to introduce a constant factor accelerating the rate of selection in the restrictive selection models discussed above.

Before going on to the more sophisticated evolutionary models in the literature, it may be useful to raise a number of questions about the proper representation of technology in such models. To begin with, the pure selection models deal with the adjustment of a disequilibrium industry state to a fixed best-practice technique. In contrast to most neo-classical models, this adjustment does take place in historical time. However, this phenomenon is not what one ordinarily considers to be technical change, which involves the continuous but perhaps uneven advance of best practice(s) themself(ves), as well as its (their) diffusion throughout the capital stock or product space of the economy. The frontier does not stand still and wait for average practice to catch up, but is rather the carrot dangling in front of the donkey (the stick presumably being backruptcy). Biological models of evolution are more justified in abstracting from this fact because mutation rates in general are considerably smaller than selection rates. This does not appear to be true for technological evolution in the present age. The neglect of this dimension is probably responsible for another widespread misconception: that disparities in average unit costs correspond to different *techniques* of production. This confusion can be dispelled somewhat by reexamining the vintage models which experienced a brief vogue in the 1960s (the original sources are Kaldor and Mirrlees, 1962, Salter, 1962, and Solow, 1960). These models had the virtue of recognising that (a) best practice is subject to continuous change, and (b) under the embodiment hypothesis the capital stock will always be a composite of investment slices acquired in the past (the vintages) and unit costs will reflect this fact in some way (most simply as an average over the vintages). Thus differences in unit costs need not correspond to differences in choice of technique. They could also be due to different time profiles of vintages of the 'same' technology. Thus a better distinction than that between a static best-practice frontier on the one hand and static techniques on the other is between vintages of a given technology or *technological trajectory* (cf. Dosi, 1982, 1984; Nelson and Winter, 1982, pp. 258–62; Silverberg, 1984) and changes of trajectory. A perusal of the management literature reveals that businessmen are well aware that capital equipment is continually being improved, so that, given that they are already established on a technological trajectory, it is very unlikely that they will be investing very far behind best practice (the problem of technologies with pronounced dynamic returns to scale is perhaps an exception here). The problem of diffusion becomes significant when a real choice of technique opens up. The vintage perspective also makes clear that there is an important distinction between diffusion in current investment and diffusion in the capital stock. Sahal (1981) is one of the few authors to present data substantiating this point. The former is necessarily more rapid than the latter. Moreover, the rate of diffusion through the capital stock is only partly determined by the speed with which entrepreneurs embrace a new technology; it will remain limited even if all entrepreneurs instantaneously shift their investment to the new technology, and it is not a sign in itself of some kind of technological inertia or irrationality. Finally, the embodiment hypothesis has the advantage of establishing a connection between the rate of change of average productivity and the composition and level of investment. In contrast to the steady-state vintage models extensively analysed in the 1960s and 1970s in which the rates of growth of best-practice and average-practice productivities are identical and independent of the rate of investment, this is no longer true with disequilibrium dynamics, as has been pointed out by Clark (1980).

The fact that best-practice technology cannot be taken as fixed during the selection process requires an extension of the basic model to include the mechanism by which this frontier is expanded and explored (not to mention how expectations about this frontier influence the scrapping decision). The simplest assumption is that the frontier moves at an exogenously given growth rate which is more or less correctly anticipated by firms. The task becomes considerably more difficult when this rate is made partly endogenous, and when switch points arise between technological trajectories. Although it is clear that both (costly) search and imitation effects are at work here, there is still no definitive agreement about how they should be incorporated into a model, and this will undoubtedly remain one of the most difficult aspects of evolutionary modelling. The approach adopted by neo-classical theory—optimal innovation as a problem of maximisation or a two-period game—is certainly very much at odds with the perspective inherent in the evolutionary framework, which, as we shall see, hinges in an essential way on the stochastic nature of search processes, the problem of decision-making under irreducible uncertainty, and collective effects.

The most ambitious attempt to incorporate these features is Nelson and Winter's evolutionary model of economic growth (Nelson and Winter 1974, 1982, Chapter 9; Nelson, Winter and Schuette 1976). Here we will restrict our discussion to the salient mathematical and economic features of the basic model to the exclusion of its extensions dealing with such questions as the so-called Schumpeterian hypothesis on industrial concentration and innovation. The model is formulated in firm space, which allows the explicit treatment of diverse firm strategies with respect to technological

innovation and imitation. In contrast to the previous models, however, a deterministic mathematical formulation must give place to a stochastic computer implementation. Technologies are once again identified with the technical coefficients of linear production functions. Technical change is disembodied, however, so that, although firms may select new techniques either by copying competitors or exploring technology space around their current technique through R & D themselves, the changeover of their entire capital stock once a new technique is found is both instantaneous and costless. Investment serves to expand capacity of whatever technique is currently in use, scrapping is a stochastically varied percentage (with mean of 4 per cent) of capacity and is thus independent of technical change. The sophistication of the model resides in the stochastic rules for search and imitation, which, however, somewhat arbitrarily are only invoked if the firm falls below a threshold rate of return on capital. Thus technical change does not constitute a routine part of a firm's strategy but rather reflects dissatisfaction with its performance. Given that a firm is looking for a new technique, it will attempt to find one via either search or imitation with probabilities that can be varied between runs. Potential technologies are represented by a random array of points in the space of the logs of the capital and labour coefficients. A metric can be imposed on this space with directional weights representing possible capital or labour-savings biases. The probability of discovering a technique via local search is then inversely proportional to this distance. The probability of uncovering a technique via imitation, as in most of the diffusion theories, is proportional to the share of that technique in total capacity. Once a new technique is found via either procedure it is subject to a profitability test at current factor prices (subject to an additional stochastic error) before adaptation. A uniform price level is assumed so that profits are a function of unit costs. Wages are determined by a labour supply curve with a possible exponential time shift. Firms reinvest their profits in capacity expansion (after deducting a rate of required dividend payment). Production is always at full capacity utilisation. This results in the familiar selection structure via differential firm growth rates. Entry of new firms is also provided for using a stochastic criterion, but appears to play a secondary role in the runs actually explored.

Thus the basic structure of the model is that of a Markov process (with time-dependent transition probabilities in the event of a shift parameter in the labour supply function). Each particular computer run (for given values of the variable parameters) is a realisation of a possible economic history. Generalisations are possible by a Monte Carlo method: data from a sufficiently large number of runs can be accumulated and evaluated to see what relatively stable properties can be identified. Patterns similar to the aggregate time series data for the American economy which Solow employed in his original growth model can be produced. If the data generated by the model are subjected to a Cobb–Douglas production-function-fitting exercise, typical R^2s of 0.99 result (a fact which is less a characteristic of the model than of the mathematical vacuity of the Cobb–Douglas procedure).

The fact that the model is formulated as a Markov process does not rule out an analytical treatment, however. In an insufficiently known article by Jimenez Montaño and Ebeling (1980), the Nelson and Winter model is recast in technology instead of firm space. A differential equation for the probability distribution (representing the probability of finding any particular distribution of technologies at a given time) can then be formulated, the so-called master equation, from the individual transition probabilities for self-reproduction of a technology (reinvestment), depreciation, and adoption of a new technology due to R & D or imitation. Although the economic interpretation of the coefficients entering into some of the transition probabilities in the paper is not always entirely clear, a formula for the mean values of the frequencies of the different technologies can be derived which closely resembles the Fisher–Eigen equation for mutation/selection in biology, modified by the addition of an imitation term. A key role is played by threshold viability values governing the rise and decline of technologies. Although this review has not emphasised explicit stochastic modelling, the first part of Jimenez Montaño and Ebeling's paper also provides a good example of the kind of insight obtainable with these methods. There they formulate a pure selection/diffusion model as a Markov process. Going to mean values the well-known logistic substitution curves can be derived. One can also ask what will be the probability that a new technique initially present as a very small proportion of total capacity will become extinct after a certain period due to stochastic fluctuations, even if it is technologically superior. They demonstrate that a technique must be superior by at least a certain factor to have a good chance of avoiding this fate.

Iwai (1984b) is also an attempt to combine a selection model with a continually advancing technological frontier. Firms experience differential growth in their share of total capacity depending on whether their unit costs are above or below the industry average (the assumed relationship is that growth rates of capacity shares are proportional to the difference between the logs of costs). The first version of the model assumes a uniform price level. If oligopolistic mark-up pricing is used, Iwai makes additional assumptions about growth rates that allow the formulation to go through. Best-practice productivity is assumed to be growing at an exponential rate. Technical progress is disembodied, so that firms can jump between the unit costs associated with different techniques without any investment. Per unit time each firm has a certain probability of innovating and adopting the unique best-practice technique. Imitation is modelled as a probability per unit time of adopting a currently employed, lower-cost technique proportional to the share of that technique in total capacity. Iwai then proceeds to examine the long-run stationary distribution of firm size resulting from the combined action of these mechanisms. The main conclusion is that a plot of firm size vs. efficiency derived from this stationary distribution misleadingly suggests that over certain ranges decreasing and increasing economies of scale are operating.

Silverberg's study (1987) is an attempt to establish a basic dynamic structure governing prices and quantities in an industry driven by Schumpeterian competition, embodied and ongoing technical progress, decision rules modelled on actual business practice and reflecting the crucial role of forward-looking expectations, and taking into account certain stylised facts of industrial development. Kaldor (1983, 1985) in particular has singled out the following observations as being in basic contradiction to received wisdom and demanding a fundamental reinterpretation of the process of industrial competition and evolution:

1. Markets do not always clear in the Walrasian sense. Businessmen take this into account by carrying inventories and order books and responding to quantity signals.
2. The presence of business goodwill, differentiated products and market inertia excludes the existence of market-clearing equilibrium (and uniform) prices and mandates an oligopolistic and dynamic reformulation of the price/quantity relationship.
3. Mark-up pricing seems to be the pervasive rule of thumb in industry and trade. However, the existence of considerable variance in unit costs of firms in the same industry indicates that either there is no *tendency* to a uniform price or that competition enforces some pattern of deviation from strict mark-up pricing. Evidently some mechanism in between these two extremes must be in operation.
4. Okun's law—that the short-run elasticity of product with respect to employment is greater than one—was interpreted by Okun to imply that the short-run average cost curves of firms were declining due to the existence of overhead labour. While this is undoubtedly true, the other half of the story implies that short-run changes in demand cannot preferentially affect only marginal firms but must be more or less equally distributed over the entire industry. This is consistent with stylised facts 1–3 above about imperfect competition. Moreover, it necessitates enlarging the disequilibrium industry concept to variations in the rate of capacity utilisation of all firms instead of just the technologically marginal ones if short-period effective demand dynamics are to be integrated into a theory of economic evolution. (This connection was first pointed out by the German economist Rüstow as early as 1926; cf. Rüstow, 1951, 1984). The presence of static economies of scale also has implications for the investment decisions of firms.
5. Dynamic models are characterised by cumulative causation, i.e. negative and positive feedback loops. If competitive economies, at the industry, national and international levels, are characterised by diverse strategies and capacities, then virtuous and vicious cycles, i.e. cumulative winners and losers, should also be possible under certain circumstances.

To combine these features into a single industry-level model, a separation is introduced between the evolutionary process at the market and at the firm level. Firms' market shares in real orders are subject to a selection mechanism based on disparities in competitiveness as they are perceived at the market level, i.e. in terms of relative prices, delivery delays, possible quality factors and advertising (only the first two are explicitly incorporated in the model at this stage). In essence this is a dynamical description of the kind of 'imperfect' competition connecting price and other signals to quantity variations outlined above. The value of a single parameter encompasses the gamut from 'pure', i.e. instantaneous, competition to pure monopoly.

The cost, price, production, capacity and delivery delay variables for the individual firm change over time in response to its investment strategy, routine behavioural rules of thumb, and its success in the market. Embodied technical progress is represented as a vintage structure of each firm's capital stock. At each point in time firms have to decide how much new equipment to acquire (gross investment) and how much oldest vintage equipment to scrap, or, equivalently, the levels of net expansion of capacity (net investment) and replacement/modernisation investment. Unit prime cost is then an average over all vintages (as are overhead costs). It can be shown that it changes over time as a function of gross investment, scrapping and the differences between best practice and marginal vintage and best practice and average unit costs. The strategic parameters determining these components of investment are the desired payback period for replacement (which itself reflects expectations about the long-term rate of technical progress) and the firm's expected rate of growth of orders for capacity expansion (modified by its rate of capacity utilisation).

The routine decision rules are similar to those employed in the systems dynamics modelling tradition. Firms adjust their production level to maintain a desired delivery delay. Prices are adjusted to costs via a mark-up, but a concession must be made to the firm's competitiveness relative to the industry average (another example of a self-organisational relationship). Given a time path for the growth of best-practice productivity (which may but need not be taken as exponential), it is possible to test different investment strategies against each other under different financial regimes (e.g. internal financing from cashflow and liquid reserves vs. unlimited borrowing). In particular, an optimal payback rule can be established; firms which deviate from it are progressively driven off the market.

Silverberg, Dosi and Orsenigo (1988) extend this basic model by considering a change of technological trajectory: at a certain point in time firms are able to choose between two qualitatively different technologies, which themselves continue to evolve over time (in the standard vintage formulation investment is always in a unique best practice). In contrast to standard diffusion models, however, the adoption decision is not modelled as one of information dissemination or involving a distribution of unvarying firm characteristics. Rather, a specific skill level internal to the firm is associated with each technological trajectory which grows according to a learning-by-using rule as a function of the firm's cumulative production on the trajectory and eventually saturates. This introduces a strong non-linear positive feedback into firm productivity dynamics. In addition, the model includes an externality in the form of a public skill level available to all firms which lags behind the growth of the average internal skill level present in the industry. Realised productivity on a trajectory is the product of the underlying embodied productivity of the vintage in question and the skill level internal to the firm specific to it. This particular combination of exogenous embodied technical progress and disembodied learning makes the actual course of productivity growth even more a function of endogenous and costly investment and production effort, but also introduces a strong element of cumulativeness.

Firms' strategies are now parameterised by an anticipation factor reflecting their optimism about the development potential of the new trajectory and their ability to pre-empt their competitors by adopting earlier than their normal payback period investment criterion would allow. The other side of the coin is the real possibility of free-rider effects due to the externality. In fact, simulation runs show that, depending on the *dynamic appropriability* of the trajectory (the ratio of the rates of internal to public learning), the same configuration of anticipation factors over firms can lead to either first or middle adopters being the major net benefitors of the diffusion process. Very late laggards run the danger of being pushed on to a downward spiral and driven out of the market altogether. On the other hand, if insufficient variance is present in firms' strategies, a socially non-optimal outcome is possible. No adoption of the superior trajectory occurs because no firm is willing to incur the costs necessary to bring the technology to commercial maturity. This kind of model demonstrates that a complex tension can exist between individual behaviours and aggregate outcomes, which may indeed be the most interesting feature of social systems. In economics this may take the form of relational pay-offs to strategies (e.g. whether first or second adopters reap the profits of an innovation, or whether it is advantageous to bet with or against the majority).

The fact that an evolutionary system may lead to completely different but self-consistent long-run outcomes depending on initial conditions and/or small random disturbances has been termed 'hyperselection' (see Ebeling and Feistel, 1982, Chapter 7) in deterministic models and 'path-dependency' in stochastic models (see Arthur's chapter in this book). Hyperselection can occur if growth rates of species or strategies are non-linearly coupled. In biology hyperselection is believed to be responsible for the almost exclusive predominance of 'left-handed' organic molecules, although *a priori* both 'handednesses' are equally viable, as well as the uniqueness of the genetic code (in a more physical context this phenomenon is also referred to as symmetry-breaking). This is an example of a 'once-and-for-all' selection process which does not admit subsequent evolutionary change and locks the system into a particular structure. David (1985) has argued that the adoption of the QWERTY typewriter keyboard is an example of this kind of lock-in process in the technological realm (mostly due to learning externalities). Hyperselective/path-dependent behaviour need not rule out all further evolutionary progress, however, and can also result in the coexistence of species which have specialised and occupy different ecological niches, in each of which further evolution is possible (cf. Ebeling and Feistel, 1982, pp. 235–8). In the following section on the self-organisation of behaviour we will come back to the question of hyperselection and collective effects and their implications for political economy and economic theory.

The self-organisation of economic behaviour

The behaviour of human beings, to the extent that it is not genetically fixed for all times, is distinguished by the fact that it is *social* and subject to learning. Thus both socially inherited norms, behavioural patterns, institutions and values, as well as the ongoing interaction with other agents and the struggle for differential advantages, must be taken into account. Economic theory for the most part has detoured around this crucial fact by reducing the problem to one of static choice of the individual who is either assumed to be so small that his actions have no effect on others and their actions can be taken as given, or so overwhelmingly large that for all purposes he is the only actor. In both cases the problem reduces to one of simple maximisation in a game against nature which the 'rational' individual is at no computational loss to solve. Although the peculiar features of the intermediate case have been recognised since Cournot, von Neumann and Morgenstern, not very much of economic theory is actually based on them. Yet the intermediate case, especially once formulated dynamically, is indeed probably the only relevant one, with the other two being limiting cases of little practical interest. (In fact, even 'pure' competition, in a dynamic environment with uncertainty, ceases to be a game against nature, as the extreme instability of primary goods and spot markets and the disproportionate role of speculation in them show.)

This is especially so in non-zero-sum settings in which both competition and cooperation are potentially possible. The best-known example and the 'ideal type' to which much research has been addressed is Prisoner's Dilemma. A wide range of economic and social phenomena seem to be related to Prisoner's Dilemma, such as oligopolistic behaviour, protection vs. free trade, arms races, etc. The most widely employed static equilibrium concept—Nash equilibrium—is not Pareto optimal in this case. Other non-zero-sum games admit multiple Nash equilibria, some of which may be suboptimal. Thus the 'rationality' of a decision may depend on the predominant strategy of the other agents; the social system decides collectively. Hence the analysis of the representative agent is misplaced here. The crucial question revolves around the interaction of the collective and the individual, and individually 'rational' decisions may lead to system-level failures. This point has been illustrated in different contexts by, among others, Keynes (1936, pp. 156–7), Schwartz (1961), Schelling (1978), G. Hardin (1968), and R. Hardin (1982). It is also at the centre of the discussion of the theoretical consistency of the so-called rational expectations hypothesis (cf., for example, the contributions in Frydman and Phelps, 1983).

The key to overcoming the pessimistic implications of Prisoner's Dilemma seems to reside in going over to a dynamic context (i.e. iterated Prisoner's Dilemma) and allowing strategies to acquire a memory of their previous encounters with specific individuals. This radically changes the context of the game and permits stable cooperative strategies to emerge (cf. Axelrod, 1984). Before discussing the implications of these finding for social science modelling, we will briefly sketch some of the main ideas of what is coming to be known as evolutionary game theory, and compare it with other behavioural approaches in economics and with other possible self-organisational processes.

The idea of applying game theory to problems of evolutionary biology goes back to Maynard Smith (see Maynard Smith, 1982, for a survey of the biological literature). The argument in a sense is the reverse of that applied in Silverberg (1984). One asks under what circumstances a given distribution of genetically transmitted behavioural strategies in an animal population cannot be invaded by a mutant. The underlying evolutionary framework presupposes that animals interact pairwise and randomly and that the pay-offs of their interactions (e.g. in mating or territoriality contests) are reflected in their reproductive fitnesses. Such a distribution of strategies in the population (which can also be interpreted as a single mixed strategy with corresponding probabilities of being played) is called an *Evolutionary Stable Strategy* (ESS), and a simple, static criterion can be specified for the existence of an ESS. An ESS is a Nash equilibrium, but the converse need not be true. An evolutionary game may possess several ESSs or none at all. Thomas (1984, Chapter 8) presents a good textbook-level introduction to the subject. The ESS concept can be used to demonstrate why stags rarely injure each other in mating contests, for example, although myopically it might seem in the interests of an individual animal to be prepared to fight to the death. In iterated Prisoner's Dilemma the pay-off of an encounter of two strategies is a discounted sum of the pay-offs of the (potentially) infinite number of plays, the discount factor representing the probability of a further play (see Axelrod, 1984, Appendix B, for details on what follows). A strategy is a (possibly stochastic) rule for deciding whether to cooperate or defect on the next play given the history of the encounter until that time. There is no universally superior strategy, and a large number of pure strategy ESSs can exist depending on the value of the discount parameter. Thus constant defection is an ESS, i.e. if it is practised by all members of a population, a small number of deviants cannot invade it. But TIT FOR TAT (cooperate on the first move, do what your opponent did on the last move) is also an ESS for a sufficiently high probability of renewed encounters in the future. Thus being an ESS is a collective property, and a system may admit many mutually exclusive regimes. The question then arises of how a transition between ESSs could come about, for example, from a world of pure defection to a world of first and faithful cooperators (but quick yet forgiving retaliators) employing TIT FOR TAT. Axelrod demonstrates that if deviants enter the system in clusters instead of individually and thus have a higher probability of internal interaction than otherwise, a cooperative strategy can establish itself. This transition between high and low levels of cooperation is also reminiscent of Williamson's model of the dynamic bi-stability of oligopolistic markets (Williamson, 1965).

An Evolutionary Stable Strategy is a static concept, however. The evolutionary problem can be cast in a dynamic setting by making the reproduction rate of the frequency of a strategy in a population proportional to the difference between its average pay-off against the population and the average pay-off of all strategies against each other. This formulation, first introduced by Taylor and Jonker (1978), returns us to our basic evolutionary equation with a particular quadratic specification of fitness derived from the pay-off matrix. The concept of an ESS is now replaced by that of an attractor, and the methods of dynamical system theory become applicable. The main results in the literature are reviewed in Hofbauer and Sigmund (1984). Of particular interest here is Zeeman (1979), in which it is shown that point attractors exist which are not ESSs. If an ESS exists representing a mix of all strategies, then it is a global attractor in the interior of the population space. A non-ESS attractor in the interior need not be global, however: the interior may also contain a basin of attraction of another point attractor on the boundary. This is analogous to the phenomenon of hyperselection referred to before. The system may converge to either a coexistence of strategies or the complete elimination of some of them depending on initial conditions. Zeeman also examines the question of the structural stability of the game dynamics, i.e. whether the qualitative properties of the flow are robust with respect to small changes in the pay-off matrix. In another paper (Zeeman, 1981) he shows that the conclusions of a model may change when it is reformulated to make it stable. Finally, periodic solutions can also occur in evolutionary games.

As a model of the formation of patterns of social behaviour, evolutionary theory represents a considerable advance over the static individual maximisation paradigm by clearly underscoring the focusing effects mediating the interaction of individuals with the collective as a process in historical time. The representative agent can be dispensed with, and indeed equilibria may correspond to heterogeneous behavioural distributions. Moreover, the rationality postulate in its strong form need not be invoked. In its strong form this assumes not only that agents prefer higher pay-offs to lower ones, but that they are able to explore completely the pay-off matrix and coordinate their behaviours in a mutually consistent way *before* any interaction actually takes place (no out-of-equilibrium trading). This is indeed a rather superhuman assumption, since it not only places extraordinary informational and computational burdens on the individual agents but neatly abstracts from any realistic coordination mechanism (Walras' *tâtonnement* process obviously does not fit this bill). The various search and reaction mechanisms that have been modelled do not necessarily converge to the game-theoretic equilibria. The basic mechanism of evolutionary theory only requires that success differentially breeds more of the same strategy, whether this be due to market expansion via a profitability/growth feedback or additionally due to imitation and adoption of *ex post* successful strategies by less successful agents (which is a first step towards an economic rationality in real time). At the first level this provides a robust approach to the explanation of patterns of behaviour, norms, implicit forms of cooperation and the focusing value of institutions and regimes when behaviour involves frequently repeated interactions between agents committed to indefinite further play. The stability of the resulting industry patterns can be probed by introducing new strategies (representing Keynes's 'animal spirits' or Schumpeter's innovating entrepreneurs) and/or by varying parameters of the interaction.

The situation is somewhat different when an evolutionary game represents a less frequent kind of interaction and the parameters may be changing in a crucial but unpredictable way between interactions. This seems to be the case, for example, in the model analysed by Silverberg, Dosi and Orsenigo (1988). Changes of technological regime occur only intermittently and crucial parameters such as dynamic appropriability may vary between plays in unforeseeable ways. The asymptotic pay-off matrix depends in a very complicated way on the configuration of the anticipation parameters of all the other firms and is relational. Thus a firm cannot really decide to be second in because it cannot know the positions taken by others in advance. And because the same experiment will not exactly be repeated in the near future, little learning or convergence to an ESS, should one exist, can be expected. What can be learned from the past, however, is that an anticipatory position must be taken if someone else takes one. The resulting uncertainty about how best to do this makes a diversity of strategies with corresponding losses and gains almost inevitable and is what drives the process as a whole.

As we have indicated, the evolutionary paradigm differs in several fundamental ways from more classical approaches to the explanation of structure and change based on the concepts of optimisation and equilibrium. But it also sheds new light on the traditional question of system optimisation, both as a realisable process in natural systems and as a planning tool.

The evolutionary process differs from classical concepts of optimisation by mandating system *diversity* in order to work at all. Moreover, this diversity must be maintained over time if the system is to avoid stagnation and display sufficient resiliency. This diversity appears to be bought at the cost of lower overall efficiency compared to a uniformly 'optimal' structure.

However, the evolutionary procedure generally displays greater robustness, i.e. lower sensitivity to mis-specification of the optimisation problem and to fluctuations in the environment. It is often also computationally more efficient than traditional methods of numerical optimisation when implemented as a search procedure on an electronic computer (it should also be pointed out that evolution is an intrinsically parallel process and thus could be considerably accelerated numerically if it could be liberated from the serial architecture of present-day computers). The question of designing search procedures along evolutionary lines has been taken up by Holland (1975), Schwefel (1981), and Brady (1985), as well as in a number of contributions in Farmer (1986). Axelrod (1985) has applied Holland's approach to a genetic coding of iterated Prisoner's Dilemma strategies with play memory three moves deep to generate evolutionary solutions of considerable complexity and effectiveness. One of the strengths of evolutionary algorithms is their much higher probability of avoiding entrapment in local maxima.

The relationship between optimisation and evolution can also be looked at in the other direction: what if anything is being optimised by the evolutionary process occurring in natural systems? The answer is not always clear-cut and depends on the exact nature of the interactions between species in the system. Ebeling and Feistel (1982) provide a thorough discussion of this point. In some evolutionary models it can be shown that a function does indeed exist which monotonically increases over time as a result of the selection process. In other cases a function may only exist which is maximised at the asymptotic steady state but fluctuates before it gets there (i.e. overall system performance may decline before it finally improves). Notice that these 'extremal functions' are open-ended: they may be capable of even further increases if new species enter the system. In some cases, however, no such function can be defined. In physical and biological systems extremal functions may correspond to such key concepts as energy efficiency or reproductive fitness. In Silverberg (1984) a 'technological potential function' is derived which unambiguously separates superior from inferior techniques (not surprisingly, this function is primarily dependent on labour productivity). The question of the existence of extremal functions characterising the evolutionary process both at the system and at the agent level is of direct relevance to the validity of one of the standard procedures of orthodox economic modelling. It is usual to assume that the behaviour of individual agents can be described by a target function which they attempt to maximise. To the extent that they are interacting in a competitive environment which decides on their fortunes in terms of survivorship, growth and rates of return, it is not at all clear *a priori* which, if any, target function will be consistent with the resulting selection process. In fact such target functions should be derivable from the competitive process, and not the other way around. Thus the long-standing discussion of whether firms maximise profits or growth rates (whatever this may actually mean) is probably not resolvable at this level of analysis. It may ultimately turn out that the success of firms is not describable in terms of target functions at all.

The self-organisation of social structures can also be analysed without an explicit consideration of pay-offs. Weidlich and Haag (1983) develop a basic model of interacting populations on the assumption that attitudes or decisions have an objective and a socially conformist component. Thus the probability of someone changing his opinion is posited to be a function of the opinions held by others in society. In a limited economic sense this need not be at odds with individual egoism, since, as Keynes argued, the pay-off to speculative investment may depend on the ability to anticipate average opinion. Equally, it may be justified as a basic human instinct (herd tendency) or derived from an interpersonal utility function. On the basis of this assumption they formulate a continuous Markov process and show that qualitative collective changes in the stationary probability distribution can take place when fundamental parameters such as the strength of coupling between individuals are varied. The addition of feedbacks allows the model to be applied to the division of investment between expansion and rationalisation as a function of underlying technological trends and the dynamics of the investment climate. The particular specification chosen, based on a Schumpeterian alternation between an expansionary and a rationalising bias of pioneering investors, is shown to lead to a limit cycle under some parameter constellations, to a number of fixed points under others.

What are the implications of these admittedly very preliminary investigations of self-organisational dynamics in social systems? On the basis of a related modelling approach to regional economic development, Allen (1982, p. 110) concludes that:

> This changes the whole concept of modelling and of prediction. It moves away from the idea of building very precise descriptive models of the momentary state of a particular system towards that of exploring how the interacting elements of such a system may 'fold' over time, and give rise to various possible 'types' corresponding to the branches of an evolutionary tree.

There are a number of qualitative features of this kind of process which deserve to be recalled in connection with policy issues. One is that nonlinear systems may display hysteresis and threshold phenomena, so that it is not always possible to generalise from the effects of small policy actions to larger ones. The existence of collective phenomena and hyperselection also implies that we are not always free to do what we want, because the rest of the system may react in a very conservative, identity-preserving way, or in counter-intuitively disastrous ones. On the other hand, very astutely applied changes by an appropriately situated agent such as governments, social movements, etc. may be able to trigger a complete and self-propagating reorganisation of the system to an unequivocally more favourable state which may not be attainable by the agents acting routinely on their own. Thus while man is ultimately condemned to discover his own history in the process of making it, with the aid of the kind of reasoning we have tried to outline here, he may be able to catch a glimpse of the labyrinthine path he is treading between freedom and necessity and possibly avoid some of the more hopeless dead ends.

References

Alchian, A.A. (1951), 'Uncertainty, evolution and economic theory', *Journal of Political Economy*, vol. 58, pp. 211–22.

Allen, P.M. (1975), 'Darwinian evolution and a predator–prey ecology', *Bulletin of Mathematical Biology*, vol. 37, pp. 389–405.

—— (1976), 'Evolution, population dynamics and stability', *Proceedings of the National Academy of Sciences USA*, vol. 73, pp. 665–9.

—— (1982), 'Evolution, modelling and design in a complex world', *Environment and Planning B*, vol. 9, pp. 95–111.

Allen, P.M., Engelen, G. and Sanglier, M. (1984), 'Self-organising dynamic models of human systems', in E. Frehland (ed.), *Synergetics: From Microscopic to Macroscopic Order*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

Allen, P.M. and Sanglier, M. (1981), 'Urban evolution, self-organization and decisionmaking', *Environment and Planning A*, vol. 13, pp. 167–83.

Axelrod, R. (1984), *The Evolution of Cooperation*, New York, Basic Books.

—— (1985), 'The simulation of genetics and evolution', paper presented at the conference Evolutionary Theory in Biology and Economics, University of Bielefeld, 19–21 November 1985.

Boulding, K.E. (1978), *Ecodynamics: A New Theory of Societal Evolution*, Beverly Hills, London, Sage.

—— (1981), *Evolutionary Economics*, Beverly Hills, London, Sage.

Brady, R.M. (1985), 'Optimization strategies gleaned from biological evolution', *Nature*, vol. 317, pp. 804–6.

Clark, J.A. (1980), 'A model of embodied technical change and employment', *Technological Forecasting and Social Change*, vol. 16, pp. 47–65.

David, P.A. (1985), 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75, pp. 332–7.

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

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

Ebeling, W. and Feistel, R. (1982), *Physik der Selbstorganisation und Evolution*, Berlin (GDR), Akademie-Verlag.

Eigen, M. (1971), 'Self-organization of matter and the evolution of biological macromolecules', *Naturwissenschaften*, vol. 58, pp. 465–523.

Ewens, W.J. (1979), *Mathematical Population Genetics*, Berlin, Heidelberg, New York, Springer-Verlag.

Farmer, D. *et al.* (1986), *Evolution, Games and Learning: Models of Adaptation in Machines and Nature*, Proceedings of the Fifth Annual International Conference of the Center for Nonlinear Studies, *Physica 22D*/1–3, Amsterdam, North-Holland.

Fischer, E.O. and Jammernegg, W. (1986), 'Empirical investigation of a catastrophe theory extension of the Phillips curve', *Review of Economics and Statistics*, vol. 68, pp. 9–17.

Fisher, R.A. (1930), *The Genetical Theory of Natural Selection*, Oxford, Clarendon Press.

Frydman, R. and Phelps, E.S. (eds) (1983), *Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined*, Cambridge, Cambridge University Press.

Gandolfo, G. and Padoan, P.C. (1984), *A Disequilibrium Model of Real and Financial Accumulation in an Open Economy*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

Georgescu-Roegen, N. (1971), *The Entropy Law and the Economic Process*, Cambridge, Mass., Harvard University Press.

—— (1976), *Energy and Economic Myths*, New York, Pergamon.

Gibbons, M. and Metcalfe, J.S. (1988), 'Technological variety and the process of competition', in F. Arcangeli, P. David and G. Dosi (eds.), *Innovation Diffusion*, Oxford, Oxford University Press, forthcoming

Goodwin, R.M. (1951), 'The nonlinear accelerator and the persistence of business cycles', *Econometrica*, vol. 19, pp. 1–17.

—— (1967), 'A growth cycle', in C.H. Feinstein (ed.), *Socialism, Capitalism and Economic Growth*, London, Macmillan.

Haag, G., Weidlich, W. and Mensch, G.O. (1985), 'A macroeconomic potential describing structural change of the economy', *Theory and Decision*, vol. 19, pp. 279–99.

Haken, H. (1983a), *Synergetics: An Introduction*, 3rd edn., Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

—— (1983b), *Advanced Synergetics*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

Hardin, G. (1968), 'The tragedy of the commons', *Science*, vol. 162, pp. 1243–8.

Hardin, R. (1982), *Collective Action*, Baltimore, Johns Hopkins University Press.

Hirsch, M.W. and Smale, S. (1974), *Differential Equations, Dynamical Systems and Linear Algebra*, New York, Academic Press.

Hofbauer, J. and Sigmund, K. (1984), *Evolutionstheorie und dynamische Systeme*, Berlin, Hamburg, Verlag Paul Parey.

Holland, J.H. (1975), *Adaptation in Natural and Artificial Systems*, Ann Arbor, University of Michigan Press.

Iwai, K. (1984a), 'Schumpeterian dynamics: an evolutionary model of innovation and imitation', *Journal of Economic Behavior and Organization*, vol. 5, pp. 159–90.

—— (1984b), 'Schumpeterian dynamics, Part II: technological progress, firm growth and "economic selection"', *Journal of Economic Behavior and Organization*, vol. 5, pp. 321–51.

Jimenez Montaño, M.A. and Ebeling, W. (1980), 'A stochastic evolutionary model of technological change', *Collective Phenomena*, vol. 3, pp. 107–114.

Kaldor, N. (1983), 'Gemeinsamkeiten und Unterschiede in den Theorien von Keynes, Kalecki und Rüstow', *IFO-Studien*, vol. 29, pp. 1–10.

—— (1985), *Economics without Equilibrium*, Armonk, NY, M.E. Sharpe Inc.

Kaldor, N. and Mirrlees, J.A. (1962), 'A new model of economic growth', *Review of Economic Studies*, vol. 29, pp. 174–92.

Keynes, J.M. (1936), *The General Theory of Employment, Interest and Money*, London, Macmillan.

Losert, V. and Akin, E. (1983), 'Dynamics of games and genes: discrete versus continuous time', *Journal of Mathematical Biology*, vol. 17, pp. 241–51.

Lotka, A.J. (1924/1956), *Elements of Mathematical Biology*, New York, Dover.

Marchetti, C. (1983), 'Recession: ten more years to go?', *Technological Forecasting and Social Change*, vol. 24, pp. 331–42.

Marshall, A. (1890/1927), *Principles of Economics*, 8th edn., London, Macmillan.

Maynard Smith, J. (1982), *Evolutionary Game Theory*, Cambridge, Cambridge University Press.

Mensch, G., Kaasch, K. Kleinknecht, A. Schnopp, R. (1980), 'Innovation trends, and switching between full- and under-employment equilibria, 1950–1978', discussion paper IIM/dp 80–5, Berlin, International Institute of Management.

Monod, J. (1970), *Le Hassard et la Nécessité*, Paris, Seuil.

Nelson, R.R. (1968), 'A "diffusion" model of international productivity differences in manufacturing industry', *American Economic Review*, vol. 58, pp. 1218–48.

Nelson, R.R. and Winter. S.G. (1974), 'Neoclassical vs. evolutionary theories of economic growth', *Economic Journal*, vol. 84, pp. 886–905.

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

Nelson, R.R., Winter. S.G. and Schuette, H.L. (1976), 'Technical change in an evolutionary model', *Quarterly Journal of Economics*, vol. 90, pp. 90–118.

Nicolis, G. and Progogine, I. (1977), *Self-Organization in Non-Equilibrium Systems*, New York, Wiley-Interscience.

Poston, T. and Stewart, I. (1978), *Catastrophe Theory and Its Applications*, London, Pitman.

Prigogine, I. (1976), 'Order through fluctuation: self-organization and social system', in E. Jantsch and C.H. Waddington (eds.), *Evolution and Consciousness*, Reading, Mass., Addison-Wesley.

Prigogine, I., Allen, P.M. and Hermann, R. (1977), 'Long term trends and the evolution of complexity', in E. Laszlo and J. Bierman (eds.), *Goals in a Global Community*, vol. 1, New York, Pergamon.

Rüstow, H-J. (1951), *Theorie der Vollbeschäftigung in der freien Marktwirtschaft*, Tübingen, J.C.B. Mohr (Paul Siebeck).

Rüstow, H-J. (1984), *Neue Wege zur Vollbeschäftigung. Das Versagen der ökonomischen Theorie*, Berlin, Duncker & Humblot.

Sahal, D. (1981), *Patterns of Technological Innovation*, New York, Addison-Wesley.

Salter, W. (1962), *Productivity and Technical Change*, Cambridge, Cambridge University Press.

Saunders, P.T. (1980), *An Introduction to Catastrophe Theory*, Cambridge, Cambridge University Press.

Schelling, T.C. (1978), *Micromotives and Macrobehavior*, Cambridge, Mass., Harvard University Press.

Schumpeter, J. (1919), *Theorie der wirtschaftlichen Entwicklung*, English translation, *The Theory of Economic Development*, Cambridge, Mass., Harvard University Press, 1934.

—— (1947), *Capitalism, Socialism and Democracy*, New York, Harper & Row.

Schuster, P. and Sigmund, K. (1983), 'Replicator dynamics', *Journal of Theoretical Biology*, vol. 100, pp. 535–8.

Schwartz, J.T. (1961), *Theory of Money*, New York, Gordon & Breach.

Schwefel, H.P. (1981), *Numerical Optimization of Computer Models*, Chichester, John Wiley.

Sigmund, K. (1986), 'A survey of replicator equations', in J.L. Casti and A. Karlqvist (eds.), *Complexity, Language and Life: Mathematical Approaches*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

Silverberg, G. (1984), 'Embodied technical progress in a dynamic economic model: the self-organization paradigm', in R.M. Goodwin, M. Krüger and A. Vercelli (eds.), *Nonlinear Models of Fluctuating Growth*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

—— (1987), 'Technical progress, capital accumulation and effective demand: a self-organization model', paper presented at the Fifth International Conference on Mathematical Modelling, Berkeley, June 1985, in D. Batten, J. Casti and B. Johansson (eds.), *Economic Evolution and Structural Adjustment*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

Silverberg, G., Dosi, G. and Orsenigo, L. (1988), 'Innovation, diversity and diffusion: a self-organisation model', *Economic Journal*, vol. 98, forthcoming.

Soete, L. and Turner, R. (1984), 'Technology diffusion and the rate of technical change', *Economic Journal*, vol. 94, pp. 612–23.

Solow, R. (1960), 'Investment and technical progress', in K.J. Arrow, S. Karlin and P. Suppes (eds.), *Mathematical Methods in the Social Sciences*, Stanford, Stanford University Press.

Taylor, P. and Jonker, L. (1978), 'Evolutionary stable strategies and game dynamics', *Mathematical Biosciences*, vol. 40, pp. 145–56.

Thom, R. (1975), *Structural Stability and Morphogenesis*, Reading, Mass., Benjamin/Cummings.

Thomas, L.C. (1984), *Games, Theory and Application*, Chichester, Ellis Harwood Ltd.

Weidlich, W. and Haag, G. (1983), *Concepts and Methods of a Quantitative Sociology*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

Williamson, O.E. (1965), 'A dynamic theory of interfirm behavior', *Quarterly Journal of Economics*, vol. 79, pp. 579–607.

Woodcock, A.E.R. and Davis, M. (1979), *Catastrophe Theory*, London, Penguin.

Zeeman, E.C. (1974), 'On the unstable behavior of stock exchanges', *Journal of Mathematical Economics*, vol. 1, pp. 34–49.

—— (1979), 'Population dynamics from game theory', in Z. Neticki (ed.), *Global Theory of Dynamical Systems*, Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

—— (1981), 'Dynamics of the evolution of animal conflicts', *Journal of Theoretical Biology*, vol. 89, pp. 249–70.

25 The diffusion of innovation: an interpretative survey*

J. S. Metcalfe
PREST and Department of Economics, University of Manchester, Manchester

Introduction

The purpose of this chapter is to consider the diffusion of innovation from a number of different theoretical perspectives. In the process I will discuss the respective roles of equilibrium and disequilibrium methods of analysis, profitability and 'fashion' supply-side factors, and the interaction between changing technology and changing market environments. I shall not give equal weight to the topics, nor shall I attempt other than the briefest reference to the literature on the subject. My purpose is to highlight some alternative methods of analysis. Davies (1979), Stoneman (1983), and Thirtle and Ruttan (1987) provide valuable surveys of the (by now) immense literature on this topic.

For the purposes of this chapter, diffusion and structural economic change are treated as synonyms. In any study of innovation diffusion, we are concerned with the process by which new technological forms are integrated into the economy to impose changes upon its structure. Diffusion-related structural change may be considered at a number of analytical levels, from the macro development of an entire industry, to the micro level at which a new machine, or consumer good, is diffused to generate corresponding marginal changes in the behaviour of firms and individuals. Most diffusion research is conducted at the micro level, but the importance of the diffusion theme spreads far beyond any detailed concern with individual innovations. In terms of fundamentals, we are interested in diffusion phenomena as examples of economic change and development in how new technologies come to acquire economic significance, and, in the process, displace existing technologies either partially or totally.

Whatever the level at which diffusion phenomena are studied there are a number of basic issues which must be clarified. The first centres on whether diffusion is to be viewed in terms of an equilibrium or a disequilibrium process (Griliches, 1957), whether diffusion patterns reflect a sequence of shifting equilibria in which agents are fully adjusted and

informed, or whether, by contrast, they reflect a sequence of imperfectly perceived disequilibria lagging behind the development of a 'final' equilibrium position. Closely related to this dichotomy is the distinction between diffusion processes which are driven by changes in external events, and those which are driven by endogenous change from within. These are not as sharp a set of distinctions as might at first appear, since one can turn any disequilibrium model into an equilibrium equivalent and vice versa by a suitable definition of the information sets and perceptions of adopting agents. However, the distinction is critical to any understanding of the diffusion literature. A second, more fundamental issue concerns the decision-making procedures which are assumed to drive the diffusion process. Here the relevant distinction is between models which assume full information, classical rationality on the behalf of adopting agents, and models which postulate limited information, bounded rationality as the basis for decision-making. From these distinctions four classes of diffusion model can be constructed. Neo-classical theorists naturally find the matching of fully rational action with equilibrium models of analysis highly congenial. Others, this author included, find the conjunction of limited-information disequilibrium methods of analysis appealing, not least because of their more open treatment of human decision-making processes. Naturally, since I believe in the appropriateness of bounded rationality as a mode of behaviour, I also welcome diversity in our approaches to understanding diffusion. Boundaries are always interesting places to be, but boundary disputes are only occasionally illuminating, and are normally tedious and unproductive.

A third issue to be clarified involves the distinction between adoption and diffusion. Adoption analysis considers the decisions taken by agents, typically organizations such as firms, to incorporate a new technology into their activities. It is concerned with the process of decision-making, and leads to propositions linking the nature and timing of adoption decisions to specified characteristics of adopters, e.g. the size of firms, or their sociometric position within a communication network. By contrast, diffusion analysis is concerned with how the economic significance of a new technology changes over time. Economic significance may be measured in a number of ways, e.g. by the share of the market held by a product innovation or by the fraction of industry output produced with a process innovation. In this sense the analysis of diffusion is closely related to the analysis of technological substitution in which a 'new' technology displaces an 'old' technology (Linstone and Sahal, 1976; Nelson, 1968). The relation between the adoption pattern and the diffusion pattern depends upon a complex of factors, including differences in intra-firm rates of adoption, and time lags between the decision to adopt and the implementation of that decision.

The fourth issue relates to the specification of an innovation, and the environment within which adoption and diffusion take place. For it is frequently assumed that the innovation embodies a technology which does

---

*Collaboration over a number of years with M. Gibbons is gratefully acknowledged, as is collaboration with colleagues in PREST, in particular, Hugh Cameron. Work reported here has been funded within the ESRC programme on competitive performance. In preparing the final draft I have benefited greatly from comments and discussion at the Maastricht conference, and, in particular, from the incisive comments of the editor of this section, Gerald Silverberg.

not change over time, and, less frequently, that adoption and diffusion occur in an unchanging environment. Neither of these assumptions is helpful except in special cases of 'minor' innovations. More typically, an innovation is one step in a sequence of innovations, within a particular technological regime. These post-innovative improvements play a vital role in increasing the rate of diffusion within existing applications, and extending the technology to new applications (Georghiou *et al.*, 1986; Hunter, 1949; Rosenberg, 1982). While some post-innovative improvements may be traced back to exogenous changes in knowledge, many arise from the experiences, incentives and bottlenecks which arise endogenously during the diffusion process. Furthermore, significant improvements are often induced in technologies under competitive threat from the new technology, so that the diffusion curve is shaped by the evolving pattern of competitive advantage between rival technologies—a phenomenon which is aptly named the 'sailing ship' effect (Graham, 1956). Endogenously driven changes apply not only to technology but also to its diffusion environment. When first introduced, a new technology is evaluated in terms of a price structure which is shaped by the prevailing technology. However, the increasing economic weight of the innovation reshapes this price structure in a way which may be favourable or unfavourable to the new technology. Indeed, whenever the existing technology is eliminated, the price structure will end up being conditioned entirely upon the characteristics of the new technology. In short, the environment in which competing technologies are evaluated evolves endogenously under the pressures of technological competition. This is not to deny that it may also change for exogenous reasons unrelated to the diffusion process, or that such changes (e.g. shifts in demand structures or the supply of inputs) may profoundly affect the evaluation of competing technologies.

The fifth issue relates to the relative importance of demand and supply phenomena in the diffusion process. This is brought out most sharply when one considers the question of profitability as the incentive to the adoption and diffusion of a new technology (Mansfield, 1961; Oster, 1982). But profitability to whom, the potential adopter or the potential producer, for innovations cannot be adopted unless they can be profitably produced? Indeed, any diffusion curve is the outcome of two processes: the one relating to the development of the market for the technology, and the other relating to the creation of the capacity to supply that market. Moreover, it is the relative profitability of competing technologies, not their absolute profitability, which is important: and this will change during the diffusion process under the influence of changes in the competing technologies and in their diffusion environment. Even in that special case where sufficient capacity already exists to meet the maximum rate of demand for an innovation, supply factors expressed in terms of pricing policy and rate of output decisions cannot be ignored (Stoneman and Ireland, 1983).

The final issue to be raised in this introductory section is perhaps the most difficult of all to treat. It is the distinction between technology as knowledge and technology as artefact (Layton, 1974). The studies of diffusion and adoption on which this survey is based are almost entirely about artefacts which, although they may be improving over time, are nonetheless readily identifiable. The study of the diffusion and adoption of technological knowledge raises quite different issues, many of which are identical to those faced in the technology transfer literature: issues relating to the cognitive and assimilative capacities of different organizations, of distinguishing organizational knowledge from technological knowledge, of imperfect property rights and the appropriability of knowledge, of the balance between codifiable, public knowledge and tacit, firm specific knowledge, and of the structure of learning activity in a given technological area. Suffice to say that these issues demand a survey of their own and will not be treated further here.

In the rest of the chapter I propose to explore some of these issues in the context of three principal themes: equilibrium models of diffusion; selection dynamics and diffusion; and a density-dependent model of diffusion and technological substitution.

The section on selection and diffusion also reflects a concern with new frameworks for analysing the process of economic change. For here we are exploring the link between variety and economic change and the role of mechanisms which enhance or diminish economic variety. There are important links here with the literature on biological and ecological change (Monod, 1963) and with the literature on evolutionary economics explored elsewhere in the volume (Silverberg). These last two sections have perhaps a modest claim to follow Marshall's dictum that 'biology is the mecca of the economist'.

Equilibrium approaches to adoption and diffusion

We begin with the equilibrium approach to adoption behaviour and diffusion, in which rational adopters possess full information about an innovation at all points along the diffusion path. The approach leads to diffusion paths, in which the timing of adoption is entirely explained by objective changes in the profitability of using a new technology. Lack of information or understanding does not constrain diffusion, and contagion effects and bandwagon effects are ruled out, *a priori*, as sources of information and influence upon adopter perceptions. The essential features of this approach are: (i) the dependence of diffusion patterns upon heterogeneity of adopter characteristics; and (ii) the identity between the objective benefits of adoption and the perceived benefits of adoption.

The general structure of an equilibrium model for the adoption of a capital-good innovation may be set out as follows. There is a given population of firms for which the capital-good innovation is technologically relevant. The size of this population may or may not be changing over time. Firms differ with respect to at least one characteristic which influences the profitability (benefit) of adoption, and the profitability of investing in the innovation is consequently distributed across the relevant population according to a density function such as B–B in Figure 25.1. For our purposes, benefits can be defined as the expected present value of profits arising from the adoption of the new technology. Obviously the shape of the benefit distribution also depends upon economic characteristics of the adoption environment, such as relative factor prices and the degree of competition faced by adopters.

Immediately one can divide the population of firms into two categories; non-adopters for whom $b_i < 0$, those firms to the left of the origin who will never rationally adopt unless subsidized to do so, and potential adopters for whom $b_i > 0$. In addition to the benefit function there is a distribution across firms of the costs of adopting the innovation. For ease of exposition, we shall assume that the present value of adoption costs is the same for all firms, at the level indicated by C–C.

The proportion of actual adopter firms follows immediately as the shaded area to the right of C–C. No rational profit-maximizing firm in this section of the distribution will fail to adopt. It follows that, to generate an adoption path over time, one or both of two events must happen: either the cost of adoption falls and C–C shifts to the left; or the benefit distribution shifts to the right in a more or less uniform manner. Either way, the outcome is to increase the size of the shaded adoption area to the right of C–C. Factors determining these rates of change would include: exogenous variations in the economic environment, e.g. a change in wage levels or in the price of primary material inputs; technical improvements to the innovation in question; and developments in complementary and competing technologies. The crucial point here is that the information sets of the population of potential adopters are independent of the number of actual adopters of the innovation. They may well change for other reasons but these are unrelated to the adoption process *per se*.

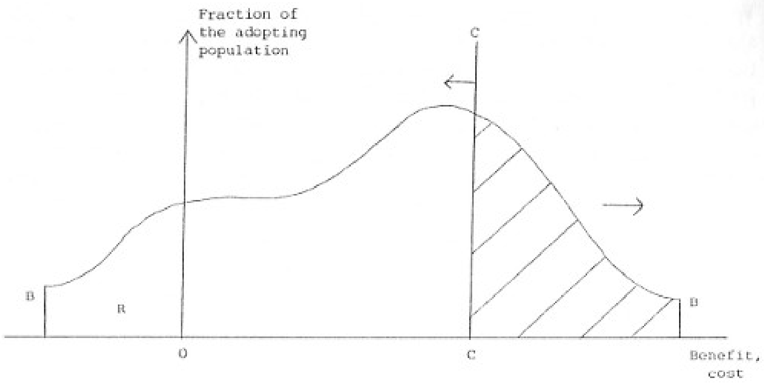


Figure 25.1

To link the adoption curve with the diffusion curve (fraction of output produced with the innovation) is not difficult in principle but obviously the shape of this latter curve will depend upon the shape of the benefit distribution, the movements over time in it and the cost distribution, and the distribution of industry output across the adopting firms. In some cases, e.g. with economies of scale in adoption, there will be an obvious correlation between the benefit distribution and the distribution of output among adopters. Unfortunately, there appears to be no research on the way in which various relevant factors influence the entire benefit distribution, a matter crucial to the shape of the diffusion curve. We do, however, have some useful insights and empirical evidence on the general sorts of issues involved. Three may be mentioned here.

The first concerns the question of interrelatedness, the dependence of the benefits from adoption upon the firm-specific environment in which the innovation is to operate. A new capital good typically has to be operated in conjunction with the existing equipment of the firm, and if the latter must be altered in any way to accommodate the innovation, the additional costs of adjustment must be added to the capital cost of the innovation (Frankel, 1955; Rosseger, 1979; Lazonick, 1985). In this way interrelatedness limits the scope for adoption. Interrelatedness factors should not, however, be limited to physical effects alone. Account should also be taken of interrelatedness between an innovation and existing labour and management skills and their organizational context, and between an innovation and the composition of the adopter's output (Feller, 1966; Rosseger, 1974).

The second point to be emphasized is the distinction between investment decisions which expand the firm's capacity and those which replace existing capacity. As far as capacity expansion is concerned, the adoption decision should be based on a comparison of the net advantages of investing in competing, newly produced capital goods, of which the innovation is only one. If the innovation offers the highest net present value it would be chosen. Old and new compete in this context on equal terms. For the replacement decision this is not the case. The definition of capital costs relevant to the existing installed technique is based not on its current reproduction cost but on actual present value in the next best alternative use. This may be as low as scrap value, net of dismantling costs, and may even be a negative sum. Hence, the old and the new do not compete on equal terms. The existing technique has costs which have been to some degree sunk. As Marshall emphasized, the nature of capital costs *ex ante* is quite different from their nature *ex post*. Indeed, this is precisely the meaning of the oft-quoted phase, 'bygones are bygones'. When the alternative capital value of the existing technique is zero, an innovation will only replace it when the savings it generates in prime costs are less than the capital charges incurred by adopting the new capital good (Salter, 1960; Lutz and Lutz, 1951). It follows from this that where replacement decisions are important (e.g. when innovations are to be adopted in a static or declining industry) then the adopter benefit distribution will reflect a number of factors including: the coexistence of different vintages of equipment in different firms; echo effects from the past history of investment decisions; and the existence of institutional factors such as second-hand markets for durable equipment.

The third and final topic to raise here concerns the optimal timing of investment (Barzel, 1968; Lutz and Lutz, 1951). Put simply, when a potential adopter is choosing between alternative, mutually exclusive investment projects it cannot be assumed that investment today is the optimal strategy. The rational investment strategy is to adopt at that date which generates the greatest net present value for the firm, and in the presence of anticipated market growth or anticipated reductions in the cost of adoption this optimal date will often lie in the future. Rational behaviour may then involve a delay in adoption, such that for all firms finding it profitable to adopt (in the shaded area in Figure 25.1) we have a corresponding distribution of their optimal adoption dates.

A number of authors have made important contributions to the equilibrium diffusion literature. Davies (1979) builds his diffusion model around the concept of adopter heterogeneity and links the shifts over time in the benefit distribution to the growth in the size of adopters and to exogenously imposed patterns of post-innovation improvement in technology. David (1975) explains the slow initial diffusion of the McCormick reaper in the American Midwest in terms of shifts in the farm size (grain acreage), distribution and economies of scale associated with mechanical reaper technology. More recently, David and Olsen (1984) have applied arguments based on the optimal timing of investment to develop a diffusion path for a durable capital good. By linking the equilibrium patterns of the diffusion to learning economies in the capital goods industry, they are able explicitly to incorporate technological expectations (Rosenberg, 1976) into the derivation of the diffusion path. Stoneman and Ireland's (1983) paper applies a similar logic to the problem but also takes account of the effects of different market structures in the supply industry upon the diffusion path.

However, our chief concern here is with the general nature of the equilibrium approach to diffusion. The key problem is not that adopters have full *a priori* information, although this is problematic enough. Rather, the problem is the assumption of information sets which are given and interpreted independently from the process of diffusion. It is this which rules out any elements of fashion or bandwagon effects as an explanation of adoption. The appraisal by firms of an innovation is complete the moment the innovation is announced. Delay in adoption can only be the result of objective circumstances, not a failure to comprehend the significance of events. In this postulated world of change entrepreneurs never fail (Sandberg, 1984).

It would appear that the central difference between equilibrium and disequilibrium models of diffusion concerns the way in which agents acquire the information relevant to their adoption decisions. In the equilibrium approach this information is given *a priori*, and if it changes it does so for reasons exogenous to the diffusion process. In the disequilibrium approach this is not so: information changes because of the diffusion of a technology, and the ways in which knowledge is acquired and the relative costs of different sources of information become crucial, if implicit, elements in the analysis. The key question in a world of costly information and limited cognitive capacity is, 'How do firms come to know the *economic* properties of a new technology?' By this is meant not simply a knowledge of the existence of the innovation in question but rather knowledge of the precise relevance of the innovation to the adopter's own particular circumstances. Broadly speaking, two types of mechanisms can generate the relevant data, remembering here that data should not be equated *simpliciter* with knowledge. The first is internal experiment, learning, appraisal and evaluation of the technology which may or may not have a formal R & D component. The second is observation of the experience of others. It is this latter which is the basis for the density-dependent diffusion process found in the disequilibrium literature. From this follows a great deal in terms of managerial and organizational behaviour, the technological sophistication of adopters, and informal and formal communication links between the population of potential adopters and between them and suppliers of the innovation (Carter and Williams, 1957; Rogers, 1983; Czepiel 1985).

Summarizing these various contributions, it is clear that they are complementary with disequilibrium models of the adoption process. It would be an error to consider them to be incompatible. The key insight they contain is not the way in which information is acquired but rather the assumed heterogeneity of adopter benefits. Introducing arguments for delayed adjustment enriches this insight; it does not diminish it (cf. Heiner's chapter in this volume).

Diffusion of innovation as a process of selection

In this section we explore some simple dynamic models of an evolutionary kind, which treat the diffusion of technology as the outcome of a process of selective competition between rival technologies. The general structure of these models is treated in the introductory chapter by Silverberg. Our concern is with the mechanisms by which a technology acquires significance, mechanisms which are based upon a sharp distinction between firm and environment. At root the argument is extremely simple. Technological variety across firms is the basis for competitive advantage. The competitive advantages of different technologies, in conjunction with certain strategic attributes of firm behaviour, determine how the rival technologies diffuse relative to one another. By following this approach one can more readily relate the study of diffusion to the study of the more general issue of competition and structural change.

Firm and environment

Technologies do not compare in the literal sense. Only firms compete, and they do so as decision-making organizations articulating a technology to achieve specific objectives within a specific environment. The outcome of their decisions is precisely what determines the economic significance of rival technologies and how this changes over time. For our purposes the firm is an organizational unit, possessing a knowledge base and a design capacity to translate that base into products and processes of production. Such a firm may be represented in terms of three attributes. First is its *efficiency*, as measured by the quality of its products and the productivity of the methods of production it employs. Efficiency depends on two interwoven aspects of the firm's knowledge base: its technological knowledge of how materials and energy are to be transformed into the desired products, and its organizational knowledge base which determines the firm's managerial capacity to plan, coordinate, control and monitor its productive activities. The second attribute is the firm's *propensity to accumulate* the ability to translate profits into the expansion of the capacity to produce its current range of products. Accumulation is a question of the perception of growth opportunities, the ability to command internal and external capital funds, the investment requirements to expand capacity, the ability to manage growth without sacrificing efficiency, and, last but not least, the willingness to expand. A firm which does not wish to grow will, by definition, have a zero propensity to accumulate. Finally, we have the *creativity* of the firm: the ability to advance product and process technology either through improvements within existing design configurations, or by the addition of new design configurations to the technological portfolio. As with the other attributes, the creativity of a firm will depend on a number of considerations including: the richness of the firm's technological environment; the resources which the firm can marshall for research design and development activities; the incentives to advance technology, influenced in part by the scale of potential application and by the threat of competitive imitation; the ability of the firm to manage the process of acquiring new knowledge (from internal or external sources); and its ability to move from knowledge to artefact, by coordinating design and development with its production and marketing activities. In a world of bounded rationality, it is hardly surprising to find inter-firm differences in creativity. Much technological knowledge is specific to the firm and not codifiable in any ready fashion. The knowledge base builds cumulatively, in part as productive and marketing experience is acquired; and what the firm perceives as a possible development in technology is contingent upon an organizational memory which reflects the history of the firm (Pavitt, 1983). Three important points need to be stressed here concerning the relationship between creativity, knowledge base and revealed performance. Firstly, it is quite inappropriate either to depict the process of creativity as one in which the organization fishes without restriction in a common pool of public knowledge, or to equate data with knowledge. Data are fragmentary, knowledge is holistic, and two firms will often interpret the same data in quite different ways. Bounded rationality will see to that. Secondly, the knowledge base of the firm is structured within the organization and develops an inertia of its own. Creativity is canalized along trajectories of advance which are self-reinforcing and become embodied in the memory and decision-making style of the organization. An established pattern of advance often degenerates with time into the only possible line of advance. The knowledge base is not infinitely malleable. It will develop in certain directions but the firm will face great difficulty in moving into non-complementary lines of production which do not draw upon established knowledge (Richardson, 1972): hence the well-documented difficulties of technology transfer between organizations. Thirdly, because firms differ in their creativity they necessarily come to differ in one or more dimensions of their efficiency. It is these differences in creativity which ultimately underpin the competitive process. They generate and regenerate the variety without which competition cannot operate.

It is appropriate to say a little more about the boundaries of the firm. For in modern conditions, the firm is typically a multi-product, multi-process operation, commanding several different technologies. We shall define the 'firm' more narrowly, as that organizational sub-unit charged with articulating a distinct knowledge base and the related design configurations, accepting that this may involve the production of several products. It follows that part of the organizational environment of the firm is formed by the umbrella organization of which it is a part. How the firm competes for capital and other resources with other sub-units and how R&D activity is distributed between the umbrella organization and firm are important determinants of accumulation and creativity.

With this clarification of the boundaries of the firm, let us turn to the question of the environment. Again, this is not a simple concept. For the moment, concentrate on the notion of the market environment to which the firm sells its products and from which it acquires productive inputs. The central point about the environment is that it is an operator, evaluating the firm's current efficiency by translating product quality into a price premium, and resource productivity into unit costs of production. This economic evaluation forms the basis of a process of competitive selection across technologies. Here one must make a number of important distinctions. First, the market selection environment may be tranquil or turbulent. A tranquil environment grows steadily with gradually evolving patterns of relative input costs and product price premium. A turbulent environment experiences discrete shocks, e.g. non-trivial, unanticipated shifts in the scale of demand and in the supply conditions of productive inputs. Important contributions to the organizational literature are concerned with how the firm should be structured to deal with tranquil and turbulent environments (Burns and Stalker, 1961; Lawrence and Lorsch, 1967). Secondly, the market environment will have a particular frequency with which the process of selection operates. A firm which sells its products or buys inputs through long-term contracts will be in quite a different position from the firm which must trade on a daily basis. Moreover, the market may operate continuously, or at discrete intervals (as it does, for example, for major civil engineering or defence contracts). Thirdly, one can distinguish market environments according to their selective force, the degree to which they punish deviations in a firm's performance from the appropriate industry average. The classic competitive market is, in these terms, one of maximum selective ferocity. By contrast, a monopolist operates in an environment of zero selective ferocity. Finally, one can distinguish market environments according to the uniformity with which they select across different firms. In part, this is a question of the segmentation of the market into non-competing groups of buyers, but it also reflects factors, such as goodwill, which imply that selection is focused more upon some firms than upon others. Stinchcombe's (1965) concept of the 'liability of newness' is a relevant example of non-uniformity in the selection environment. Now the implication of these environmental attributes is that, in conjunction with variety in the attributes of firms, they determine the differential rates of diffusion of competing technologies. To explore this more fully we need first to present a distributional view of technology.

Technology as a distribution

It is commonplace to observe that in any industry one finds not uniformity but variety of economic performance. Take, for example, unit costs of production. Instead of one common cost level, one typically finds firms with a wide variation of unit costs of producing more or less identical commodities; from best practice through average practice to worst practice. Indeed, it was precisely such an observation which prompted Salter's (1960) pathbreaking analysis of the coexistence of competing production methods in terms of vintage investment and replacement decisions. As Silverberg (1985) has shown, the efficiency of a firm will be a function of the mix of capital vintages that it operates, with this mix being changed continually under the pressure of scrapping and investment decisions. Exactly the same observation underspins studies of the frontier production function (Farrell, 1957).

Consider an industry producing a more or less homogeneous product. At any one time, the variety within it can be expressed (as in Figure 25.2) in terms of a density function showing the fraction of production, s_i, accounted for by each different unit cost level, or equivalently by the cumulative distribution function (Figure 25.3). The explanation of this variety may be found in one or more of the following factors: the environment is not uniform so that different firms pay different prices for identical productive services; the firms differ in their ability to organize the process of production; or the firms are employing technologically distinct methods of production or different combinations of the alternative technological vintages. Since our primary concern is with the diffusion of technologies, we shall equate, for pedagogic purposes only, differences in unit costs entirely with technological differences in the methods of production.

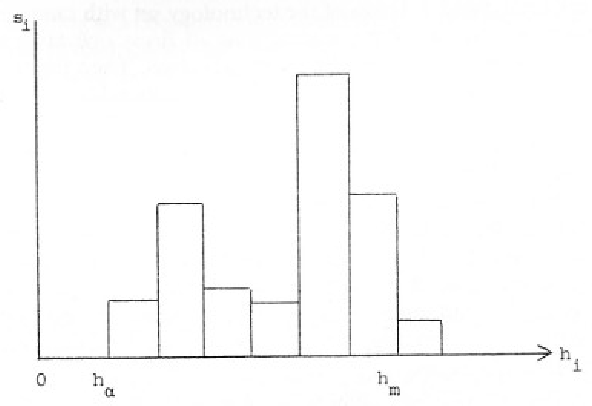


Figure 25.2

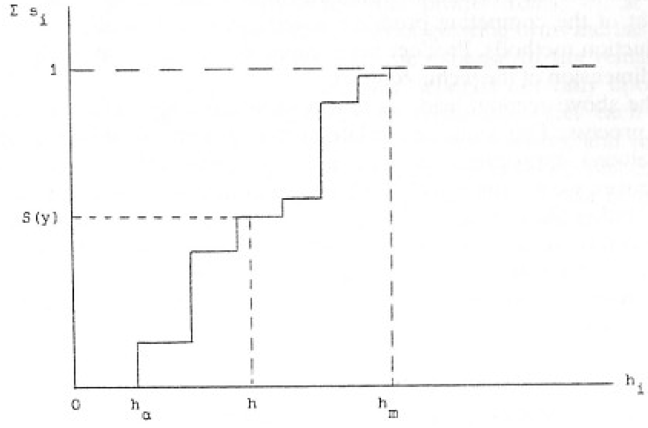


Figure 25.3

Variety is then defined in terms of the technology set with range $h_\alpha - h_m$. It helps exposition further if we assume that all firms operating the same process technology do so with the same unit costs. Then the fractions of production accounted for by each different technology, their market shares, accurately measure the different degrees to which they are currently diffused. Over time this distribution changes under the influence of three sets of forces, process innovations and related post-innovation improvements, process imitations, and market selection. The first two redefine the elements of the technology set and are properly related to matters of creativity. The third changes the economic weight of the rival technologies and depends on the environmentally contingent interaction between efficiency and fitness. It is this selection-induced respositioning of the elements in the technology set which traces out the diffusion curves of the different technologies.

This argument applies equally to questions of variety in product qualities. Each product may be considered as a bundle of characteristics. Depending on the (implicit) market valuations of these characteristics, the different products will command different market prices. In a uniform environment, the pressure of selection continuously changes the economic weight of the competing products exactly as it does with respect to the production methods. Product innovation and imitation similarly redefine this dimension of the technology set.

The above account leads us to a multi-technology picture of the diffusion process. The traditional bilateral model (new vs. old technology) is not always appropriate as a basis for analysing diffusion phenomena. Moreover, we see the direct link between diffusion and competitive selection. Rather than being on the fringes of competitive analysis, diffusion is quite central to it. Diffusion and competition are based upon rivalry and variety, and no dimension of variety is more telling in the long run than that which is based upon efficiency differences and canalized patterns of technological change.

A review of simple selection models

We can now put these general arguments to work in a sequence of illustrative but simple selection models which build upon the ideas of Steindl (1952), Downie (1955), Nelson and Winter (1982), and Iwai (1984). We consider first selection by differential accumulation in uniform environments of maximum selective ferocity. We then extend this formulation to cover imperfect customer selection, increasing returns and barriers to exit. The common theme running throughout these cameo studies is of firms earning differential quasi-rents which are then deployed to finance market expansion and creativity.

Selection by differential accumulation

Consider first a technology set comprising a given variety of process technologies for producing a homogeneous commodity. The market is tranquil, growing at a constant compound rate g_d. It is uniform and of maximum selective ferocity, so that every firm is forced to sell its output at the common market price, p. At each point in time, the market clearing value of p depends on the market demand curve and the schedule of supply, itself dependent on the cumulative total of capacity currently invested in each competing process technology. All firms pay the same price for productive inputs (uniformity), and this structure of input costs does not vary during the selection process (tranquility).

Each technology is of the constant returns kind with capital: output ratio, v_i. Unit costs with each technology u_i, are the sum of a fixed component, h_i, and a further component, $\rho_i g_i$; this latter reflecting the link between the rate of growth of the firm and its ability to avoid consequential current cost penalties. For expositional purposes we will term h_i, unit costs for the ith technology. Unit cost in this sense includes the element of normal return on capital invested in that technology.

At any point in time, the ith profitable technology has market share s_i, and this is our measure of its relative degree of diffusion. Consider now how these market shares change over time. Given the product price, firms with different technologies earn differential profits (rents) which they invest in further capacity expansion. The faster growing firms increase the relative diffusion of their technologies at the expense of the remaining firms. However, how quickly a firm grows depends not only upon its efficiency but also upon its propensity to accumulate. Let each firm acquire investible funds from internal and external sources, and let the ratio of the growth rate to rate of profit for a given firm be π_i. This is also the coefficient of investment in the particular technology articulated by that firm. Combining these hypotheses we can write

$$g_i = \left(\frac{\pi_i}{v_i}\right)[p - h_i - \rho_i g_i]$$

which gives

$$g_i = f_i(p - h_i), \tag{1}$$

with

$$f_i = \frac{\pi_i}{v_i + \pi_i \rho_i} \tag{2}$$

being the firm's accumulation coefficient. The firm has a higher propensity to accumulate, the smaller is its capital:output ratio, the more effectively it manages growth (smaller is ρ_i), and the larger is its investment ratio π_i. A firm which did not wish to grow or could not manage growth would have a zero value of f_i.

We can now divide the technology set into two groups. Profitable technologies in which investment is occurring, $p > h_i$, and bankrupt technologies for which output is zero, $p < h_i$. On the borderline of survival between these groups are the marginal technologies, in which output is positive although investment has ceased, $p = h_m$. Implicit in this partition is the assumption that bankruptcy results in immediate withdrawal from the market (on this see below).

Under conditions of uniform accumulation, $f_i = f$, the multi-technology diffusion process will operate in the following way. The aggregate growth rate of the profitable firms is $g = \Sigma s_i g_i$, from which we derive

$$g = f[p - \bar{h}] \tag{3}$$

with $\bar{h} = \Sigma s_i h_i$, the level of average practice cost units. Two types of selective situation may then apply, depending on whether there exists a marginal technology. If there is, then the market price will be fixed at the level of $p = h_m$, and any discrepancies between g and g_d will be reflected in a changing degree of capacity utilization in this marginal technology. Either $g > g_d$, in which case the utilization of the marginal technology is declining, or $g < g_d$, in which case its degree of utilization is increasing. In each case the marginal technology must eventually cease to be marginal (it moves into either bankruptcy or profitability). When this happens we find that p is no longer technologically determined and rises or falls over time as $g \gtrless g_d$ until g converges on g_d to give

$$p = \frac{g_d}{f} + \bar{h} \tag{4}$$

In what follows we can, with minimal loss of content, focus solely on a situation where marginal technologies have been eliminated.

To depict diffusion phenomena in a multi-technology context we have a number of options. The first is to consider directly the market share of the ith technology, which evolves according to the relation

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = s_i(g_i - g) = f[\bar{h} - h_i] \tag{5}$$

It follows that the rate of diffusion is linked directly to that technology's distance from average-practice technology. Provided the technology has unit costs below the average level, its level of diffusion increases; otherwise it declines in relative significance, surviving as long as it remains profitable. It can be shown that this process brings to economic dominance the lowest-cost, best-practice technology, the diffusion of all other technologies dropping in relative terms to zero. Crucial to this process is the fact that this multi-technology diffusion process continually defines the value of $\bar{h}$, since the more efficient technologies are always increasing in relative weight. Thus any technology which is not best practice but which starts with lower costs than average will first experience increased diffusion until such time as $\bar{h}$ falls sufficiently to make it above average in terms of unit costs, when its level of diffusion declines. Ultimately, $\bar{h}$ will tend to the best-practice unit cost level, h_α.

A second way to depict the diffusion process is in terms of the evolution of the statistical moments of the technology disltribution. An obvious candidate is the evolution of average-practice unit costs, according to the zero order replicator equation (Schuster and Sigmund, 1983, and Silverberg in this volume).

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}t} = \sum_i \frac{\mathrm{d}s_i}{\mathrm{d}t} hi = \sum_i s_i(g_i - g)h_i = C(g, h) \tag{6}$$

The rate of change of average-practice costs is equal to the covariance between growth rates and unit cost levels in the technology set. This is a fundamental and typical result in a selection framework (Price, 1972). The evolution of any statistical movement of the technology distribution depends on some measure of variety of performance across the distribution. However, we can go further, for $[g_i - g] = f[\bar{h} - h_i]$, which gives

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}t} = -fV(h) \tag{7}$$

where $V(h)$ is the variance of unit costs across the profitable firms. This result is directly analogous to Fisher's fundamental law of natural selection in genetic biology (Nelson and Winter, 1982, p. 243). Here we also see the accumulation coefficient as a measure of the ferocity of the selection process. Competition is fiercer, and average practice falls faster, the greater the propensity to accumulate of the firms in this industry. By an exactly similar argument, one can show that the variance of unit costs evolves in proportion to the third movement around the mean of the technology distribution.

Yet a third method of depicting the diffusion process, is to consider the evolution of the cumulative distribution function of the technology set (Iwai, 1984). Choose a unit cost level, y, $h_\alpha \leq y < h_m$) and define $S(y)$ as the cumulative fraction of industry output produced by technologies with costs less than or equal to y (Figure 25.3). This choice of y divides the distribution of profitable technologies into two parts. Let $\bar{h}_1$ be average unit cost for all technologies with $h_1 \leq y$, and let h_2 be average unit cost for the remaining technologies, with $h_2 > y$. Then it can be shown that $S(y)$ evolves with time according to

$$\frac{\mathrm{d}S(y)}{\mathrm{d}t} = f[\bar{h}_2 - \bar{h}_1]S(y)(1 - S(y)) \tag{8}$$

that is, logistically, at a rate determined by the propensity to accumulate and the cost gap $\bar{h}_2 - \bar{h}_1$. Of course, $S(y)$ will not follow a logistic curve through time, since this cost gap is itself changing as selection proceeds.

Each of these methods provides a valid way of describing the multi-technology diffusion process, and each requires some minor modification once we admit the existence of marginal technologies and the transition into bankruptcy. The point here is that with the growth of capacity, the price of the product follows a halting downward trend, so squeezing progressively more technologies into bankruptcy. Ultimately, best practice dominates the production of the industry. While those technologies which can satisfy the condition $f[h_i - h_a] < g_d$ survive, they will all ultimately acquire a zero measure in terms of relative diffusion. This leads directly to some crucial distinctions. Absolute growth in the output from a technology should not be confused with its diffusion, which is a relative matter: a technology can grow in terms of absolute measures but still be declining in economic significance. Similarly, one must distinguish significance from survival. To survive, all that is required is that a technology's unit costs permit profitable production, $p > h_i$. To have lasting significance a technology must be best practice. Notice that the margin for the survival of non-best-practice technologies is greater the greater is the growth rate of the market environment. It is static conditions that impose the 'survival of the most efficient' as an equilibrium requirement.

Customer selection

We maintain the assumption of a uniform propensity to accumulate and analyse the effects of alternative specifications of the diffusion environment. We consider first the consequences of weakening the force of market selection, so that a firm with a 'high' price is not immediately punished by finding its sales drop to zero. In effect, this introduces elements of dynamic imperfect competition into the selection process. To fix ideas, imagine that customers in the market at any time are divided between the firms in a particular pattern but allow the firms to set different prices for the same product. Each firm has a different customer base but interaction between the customers of different firms spreads knowledge about the price distribution, so that customers switch gradually from high-priced to lower-priced firms. Let the customer learning mechanism reflect random mixing between the customers of different firms and follow the rule

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = \delta \sum_j s_i s_j (p_j - p_i)$$

When customers from two firms interact, the one paying the higher price switch to the lower price firm at a rate proportional to the coefficient δ. In so doing they redistribute the derived demand for the different process technologies. This particular adjustment process generates a growth rate of demand for each firm.

$$g_i = g_d + \delta(\bar{p} - p_i) \tag{8}$$

where $\bar{p} = \Sigma s_i p_i$, is the industry average price, and δ is the coefficient of customer selection. Equation (8) satisfies the necessary aggregation condition that $\Sigma s_i g_i' = g_d$. When $\delta = \infty$ we have the case already considered, effective perfect competition, with each firm forced to charge the same price. When $\delta = 0$, we have a world of independent monopolies since the customer base of each firm is quite independent of the price that it charges relative to other firms. Customers are completely loyal to their existing supplies. In between we have a world of dynamic imperfect competition—dynamic because the demand curve for each firm shifts over time in proportion to its average price differential. Combining this customer selective mechanism with the differential accumulative mechanism, and imposing the balance condition that $g_i = g_i'$, we find that

$$g_i = \frac{fg_d + \delta g}{f + \delta} + \frac{\delta f}{f + \delta}(\bar{h} - h_i)$$

If we again ignore the existence of marginal firms, and focus on those tranquil situations in which market demand and aggregate capacity grow in step, we find

$$g_i = g_d + \frac{\delta f}{f + \delta}(\bar{h} - hi) \tag{9}$$

Now (9) is a selective mechanism of identical form to (1), except that imperfect market selection interacts with accumulative propensity to slow down the rate at which selection operates. At the extremes, when $\delta = \infty$, we have market selection at its maximal rate; when $\delta = 0$, selection is ruled out. In this latter case $g_i = g_d$ for all the technologies, and the diffusion process has its structure frozen into a pattern determined by initial market shares.

Turning now to the evolution of the technology distribution, we find that

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}t} = -\frac{f\delta}{f + \delta} V(h), \tag{10}$$

which reduces to (8) when customer selection is of maximal ferocity. When $\delta = 0$, we naturally find that the technology distribution ceases to evolve. The one line of argument which cannot be carried over to these conditions of imperfect customer selection is that relating the market demand curve to a uniform price. For now we have a distribution of prices and cannot define a market demand curve. It is easily shown that $(\bar{p} - p_i)$ is proportional to $(\bar{h} - h_i)$, and hence the variance of prices is given by

$$V(p) = \left[\frac{f}{f + \delta}\right]^2 V(h) = \frac{f}{f + \delta} C(p, h) \tag{11}$$

while

$$\frac{\mathrm{d}\bar{p}}{\mathrm{d}t}=\delta^2 V(h) \tag{12}$$

so that selection generates lower average prices as well as lower average unit costs. The survival condition also requires some modification. A firm remains viable as long as it covers unit costs but the permissible lag behind average-practice costs is greater the greater is the growth rate of demand, and the lower is the ferocity of customer selection. Once again we find that rapid market growth is conducive to the survival of inefficient technologies. From this we conclude that the ferocity of the market selection environment is a key determinant of patterns of technology diffusion. The more benign the environment, the smaller is δ, and the less quickly does the best-practice technology establish its dominance. More generally, one can allow the ferocity of customer selection to be uneven across the technologies (specifying δ_{ij} different for any pair of firms) with effects similar to those generated by non-uniform fitness. A firm whose customer base is relatively 'sticky' can tolerate greater inefficiency and thus survive longer than in a world of uniformity in the market selection process.

A digression: increasing returns and the Verdoorn Law

We return again to the basic selection model but allow for increasing returns to scale as firms expand their productive capacity. Again, to see the essentials, imagine that scale effects are Harrod neutral (leave the capital:output ratio of each technology unchanged), and that the scale elasticity, ε, is the same for each technology. Then, along familiar lines, average-practice unit cost evolves according to

$$\begin{aligned}\frac{\mathrm{d}\bar{h}}{\mathrm{d}t}&=\sum_i \frac{\mathrm{d}s_i}{\mathrm{d}t}h_i+\sum_i s_i\frac{\mathrm{d}h_i}{\mathrm{d}t}\\ &=-fV(h)-\varepsilon\Sigma s_i h_i g_i\end{aligned}$$

Simplifying this expression further we can write

$$\frac{\mathrm{d}\log(\bar{h})}{\mathrm{d}t}=-\left[f(1-\varepsilon)\frac{V(h)}{\bar{h}}+\varepsilon g_d\right], \tag{13}$$

which is precisely the 'form' of relationship postulated by the Verdoorn Law, with empirical values of the scale elasticity typically found in the range $0\leq\varepsilon<1$.

Barriers to exit

It is natural for the study of technical change to focus upon the frequency and character of innovation events. We cannot, however, forget the converse phenomena, the elimination of once economic technologies and the bankruptcy mechanisms which bring this about. In fact, the processes of selection we have described depend upon two rules of behaviour: the accumulation rule, which relates capacity growth rate to profit margin; and the survival rule, which states that non-profitable technologies exit the industry.

There are, however, good reasons to expect the survival rule to be violated in many situations. Even though a firm cannot cover its costs, it may still draw upon accumulated capital resources in the hope of better times to come. If it is part of a larger multi-product organization, the profits generated elsewhere may be used to shore up the bankrupt technology. Finally, public subsidy may and often does keep 'bankrupt' technologies in operation. Now whatever the other implications of these violations they do have one obvious consequence: they slow down the selection process because the capacity invested in the bankrupt technologies remains in production and depresses the market price. Consequently, the profitable technologies are less profitable and their collective rate of diffusion is correspondingly reduced. Certainly, in any study of diffusion and the competitive process the conditions and timing of technology exit should be given due consideration.

Some extensions

It is possible to develop this simple selection model in a number of ways. Differences in product quality which generate different product prices may be combined with cost differences to give two sources of differential profitability. Similarly, different propensities to accumulate can be introduced, so that the simple link between efficiency and capacity growth rate is broken. Then it no longer follows that it is the best-practice technology which rises to relative market dominance (Metcalfe, 1984). One may also introduce the effects of turbulence in the environment, with demand curve or factor price shocks displacing the selection process from its initial path. But in each case the central logic of the multi-technology diffusion process remains unaltered. Selection works on economic variety at a pace which is greater, the greater is the economic variety at any point in time.

A Marshallian diffusion process

While the approach to multi-technology diffusion in the previous section has the advantage of generality, it has the consequential disadvantage that explicit diffusion curves cannot be derived. In this final section we outline a process of competition between an old and a new technology which permits the derivation of a solution for both the traditional diffusion curve and for the relative diffusion or substitution curve. The approach is Marshallian, both in terms of the use of partial equilibrium methods, and in illustrating the fundamental distinction between expanding and contracting industries. The latter are in a long-period situation of capacity expansion, while the former are necessarily in a short-period situation with given productive capacity. The approach also allows an explicit analysis of the role of profitability and fashion, as incentives to adoption and as stimuli to capacity expansion. Demand and supply sides of the diffusion process are brought together to determine simultaneously patterns of output, prices, unit costs and profitability for the two technologies. The approach also permits an integration of equilibrium and disequilibrium approaches to diffusion, precisely to emphasize their complementarity. This integration, however, is bought at a price, namely the reliance on path-independent methods of analysis. For we construct positions of equilibrium which are quite independent of the paths towards those long-period positions. While this method has an honourable history in both classical and neo-classical economic thought, it seems to be peculiarly inappropriate for the analysis of the process of competition and technological change (see the chapters by Dosi and Orsenigo and by Arthur). It is worth noting that the type of process outlined below has direct parallels in the ecological literature on inter-species competition. It is in fact an example of a density-dependent selection process, in which competition settles the two technologies into their respective niches.

The problem

The process of competition is between two technologies which supply the same productive service to users. We choose units such that one unit of the old commodity provides one unit of productive service, while α units of the new commodity ($\alpha < 1$) provide one unit of the productive service. The coefficient α measures the qualitative superiority of the new technology and thus its equilibrium price premium. In equilibrium the price of the old, p_o, and new, p_n, commodities must satisfy the condition $p_n/p_o = \alpha$. The two commodities also have different, constant returns technologies of production, and correspondingly different equilibrium supply curves. Within each technology, all firms produce under the same cost conditions. The market environment is uniform, and there is a given, static demand curve for the productive service. The two technologies draw upon different markets for their productive inputs, and as the output of each expands external diseconomies are encountered which increase the equilibrium supply price.

The equilibrium niche

At the date of innovation of the new technology, the old technology supplies the entire market for the productive service. This market is in long-run equilibrium with price p_o^* and output level C_o^*. The entry of the new technology redefines this position and creates a new set of niches for the two technologies. Either the old is completely eliminated, or the two technologies share the market for the productive service. The outcome depends on the demand curve for the productive service, the two supply curves, and the qualitative superiority of the new technology (α). By comparing prices and quantities in terms of units of the new material we can depict the possible outcomes in Figure 25.4.

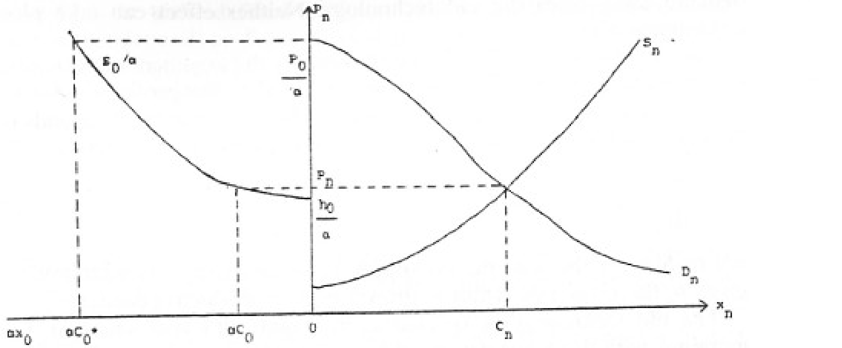


Figure 25.4

D_n is the excess demand curve for the new material, which is derived by subtracting, at each price, the supply of productive service from the old technology from the corresponding market demand for the productive service. Hence the intercept of this curve with the vertical axis has the value p_o^*/α. With S_n as the equilibrium supply curve for the new technology, its equilibrium niche is C_n and corresponding price p_n. Prior to the innovation date the old technology has output αC_o^*. In the new niche the output from the old technology has contracted to αC_o on its supply curve. The long-run demand is shared between the two technologies in the proportion s_1 given by

$$s_1 = \frac{C_n}{\alpha C_o + C_n} \tag{14}$$

In Figure 25.4, the two technologies coexist. This need not be so. If $h_n > p_o^*/\alpha$, then the new technology cannot be established, which, if $h_o/\alpha > p_n$, then the old technology certainly cannot survive. Some supply inelasticity is crucial to its survival (Harley, 1975). It readily follows that s_1 is greater, the smaller is α, and the greater is the cost superiority of the new technology.

Transition to the equilibrium niche

Consider now the process by which the long-run equilibrium position is attained. For the new technology to displace the old, its productive capacity must be built up and users must be induced to switch their demand away from the old technology. Neither effect can take place instantaneously.

On the capacity growth side, we can employ the argument of the second section of this chapter. Growth is proportional to the profitability of the new technology, with the added constraint that unit cost now depends on the scale of production from the new technology. We can write this as

$$\frac{dx_n}{dt}=fx_n[p_n-S(x_n)] \tag{15}$$

where $S(x_n)$ is the long-period supply curve of the new commodity. For given p_n this gives a determinate growth path of productive capacity.

The old technology is, of course, in a quite different situation. It is marginal, with firms just earning the normal return required to keep them in production. Investment decisions are no longer relevant, and the only decisions to be made concern the rate of contraction down the long-period supply curve.

On the demand side, users have to learn of the attributes of the new technology before they will adopt it. Following the argument in the first section, this leads to a learning process in which non-adopters learn by observing the experience of existing adopters, so creating a 'fashion' element in the diffusion process. Under conditions of random interaction, and no external influences in the learning process, this leads to the following differential equation for the growth of demand for the new commodity.

$$\frac{dx_n}{dt}=\beta x_n[D_n(p_n)-x_n] \tag{16}$$

where β is the constant, adoption coefficient and $D_n(p_n)$ is the long-run demand curve for the new commodity.

Now for a given value of p_n we have two fully determined growth paths for the new commodity. Moreover, if $D_n(p_n)$ and $S_n(x_n)$ are linear functions, then (15) and (16) reduce to a pair of logistic differential equations. But for arbitrary p_n they also generate different time paths for capacity and demand. In a closed economy this cannot be so. Capacity and demand cannot grow in an inconsistent fashion, for entrepreneurs will neither tolerate capacity shortages nor excess capacity. In the short term such deviations are probable but not in the long term. So we seek balanced paths of diffusion to generate the secular trend of the diffusion process (Kuznets, 1929; Burns, 1934), along which p_n varies to maintain the growth rate of capacity equal to the growth rate of demand.

In the particular case where $D_n(p_n)$ and $S_n(x_n)$ are replaced by first-order approximations then the balanced diffusion curve of the new technology can be derived explicitly. Along the balanced path we find that

$$\frac{dx_n}{dt}=Bx_n[C_n-x_n] \tag{17}$$

where C_n is the niche defined in Figure 25.4, and B is the diffusion coefficient which depends on β, f and the slopes of all the equilibrium curves in Figure 25.4. A higher value of f or of β increases the magnitude of β. C_n, is independent of both these dynamic coefficients (Cameron and Metcalfe, 1987).

Solving (17) leads to the familiar logistic equation

$$x_n(t)=C_n[1+A\exp(-BC_nt)]^{-1} \tag{18}$$

where A is a constant depending on the scale of output of the new technology at the innovation date. Despite the complexities of this competitive process, with endogenous changes in price, profitability, and unit costs, the output of the new technology follows a logistic curve towards its equilibrium niche. The properties of this process are more readily seen in Figure 25.5. In Figure 25.5(a) we have the curve of balanced logistic growth for X_n, and in 25.5(b) the curve of the growth rate in output of the new technology showing the familiar pattern of retardation. In Figure 25.5(c) is the path of decline in the output of the old technology, while 25.5(d) shows the path of price and output for the new technology. Given that f and β are both finite, the path starts from point a and reaches the long-period

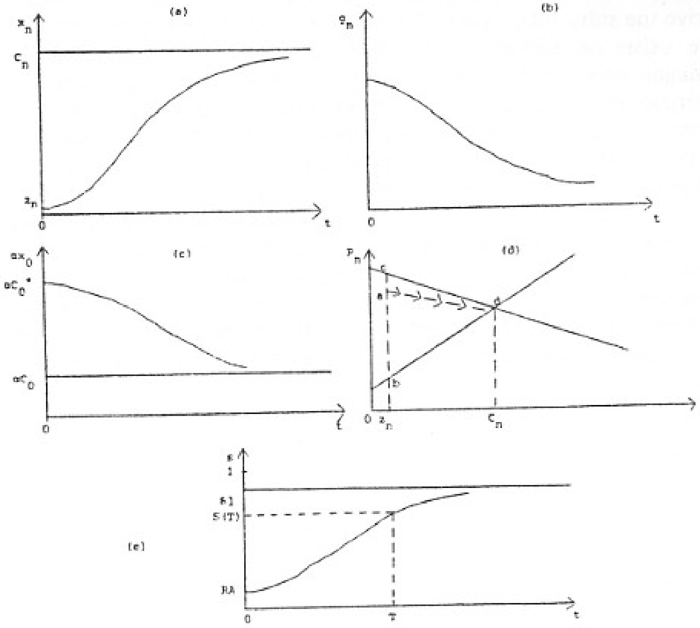


Figure 25.5

position at d. Two special cases then fall into place. With $\beta = \infty$, there is no customer learning and $p_n(t)/p_o(t) = \alpha$ throughout the diffusion process. The path followed is then along the demand curve from c to d. With $f = \infty$ we have the corresponding case of no accumulation constraints on the diffusion process, and the path followed is along the supply curve from b to d.

From this we draw an important implication. Along any balanced path with $\beta < \infty$, we find that $p_n(t)/p_o(t) < \alpha$. This means that during the diffusion process the new technology is always more profitable to adopt than the old technology, and it is only the lack of familiarity with its attributes that holds back adoption. Any adopter switching to the new commodity will lower the cost of acquiring the productive service, although this cost advantage declines over time. In the long-run position the new commodity will sell at its full premium, and only then will the marginal user be economically indifferent between the two technologies. Thus the diffusion process has built into it a clear incentive to switch to the new technology once its attributes are properly understood. Indeed, without this economic incentive, the market for the new technology could not begin to grow.

The substitution curve

Having determined the outputs of the old and new materials we may now derive the substitution curve for the new technology. An extensive literature exists on technological substitution (Linstone and Sahal, 1976; Mahajan and Peterson, 1985), but no adequate theory exists on the determinants, as distinct from the empirical properties, of substitution curves. Within this two-technology diffusion framework, the substitution curve corresponding to a balanced path can be derived as follows, although one must first decide how market shares are to be measured. The obvious method is to employ current prices and to measure shares in total expenditure, but this is unnecessarily complex and permits no easy solution since the prices are changing over time. However, since the materials are physical substitutes, we can justifiably use their relative efficiency in supplying production services to compare them on a common basis. Thus the market share of the new material at time t can be computed as

$$s(t) = \frac{x_n(t)}{\alpha x_o(t) + x_n(t)} \tag{19}$$

This is equivalent to combining the old and new commodities at their long-run relative prices.

On differentiating (19) we have

$$\frac{\mathrm{d}s(t)}{\mathrm{d}t} = s(t)(1 - s(t))(g_n(t) - g_0(t)) \tag{20}$$

Provided the growth rate of output of the new material, $g_n(t)$, exceeds the corresponding growth rate of the old, $g_o(t)$, then $s(t)$ increases over time. At what rate it increases is not transparent, since the two output growth rates are themselves varying over the substitution process. However, these growth rates are also determined once we know the balanced diffusion path. Taking account of this we find that the substitution curve is given by

$$s(t) = s_1[1 + RA\exp(-BC_n t)]^{-1} \tag{21}$$

where s_1 is as determined in (14), and R is the ratio of sizes of the equilibrium markets for the productive service in the old and new equilibrium positions. $[R = \alpha C_o^* / (\alpha C_0 + C_n) < 1]$ R is a measure of the long-run impact of the new technology upon the market for the productive service. All the remaining coefficients in (21) are as defined in (18).

Figure 25.5(e) shows the typical path of substitution, which follows a logistic curve from a value $s(0) = RA$, towards the upper asymptote s_1. It is not surprising that the coefficient BC_n determines simultaneously the rate of diffusion and the rate of substitution, since the output of the old technology is simply responding passively to the growth of the new commodity. It is the dynamics of the new technology which is driving the process.

As an index of the rate of substitution we can derive the time, T, taken to reach some target substitution level $s(T)$. Simple manipulation gives this as

$$T = \frac{1}{BC_n}\left(\log RA + \log\left[\frac{s(T)}{s_1 - s(T)}\right]\right) \tag{22}$$

This time is greater the smaller is BC_n and the greater is R. The more significant the new technology, the longer, *ceteris paribus*, will it take to reach a given level of substitution.

Thus the substitution curve is akin to the hands of a clock, tracing out the surface phenomena of substitution, while, hidden from view, a complex dynamic of prices, production and profitability drives the process of technological competition. For our purposes the significance of (21) lies in its non-arbitrary nature. The substitution curve is an explicit reflection of the process of competition between the old and the new materials as reflected in the relative profitability of using and producing them.

One advantage of this approach is that it enables us to enquire into the factors hidden behind the substitution curve and thus to investigate the effect of parameter changes upon the substitution process. Any such change can be assessed in terms of its impact on s_1, R, BC_n and A. One possibility is to compare substitution processes identical in all respects except one, and to infer the effects of the specified difference upon the substitution curve. Two examples will suffice to illustrate the method. Consider first the effects of an improved process for producing the new material, corresponding to a rightward shift in the supply curve $S_n(x_n)$.

Taking first the effects on the long-run position we see that C_n is increased and p_n reduced with the consequence that C_0 is also smaller. Together these changes mean that an improved process is associated with a higher value of s_1. With respect to the dynamics of adjustment it follows immediately that the improved process is associated with a higher value of BC_n, the substitution rate coefficient, and a higher value of product RA. Thus improved methods for producing the new material work to expand its equilibrium market share and ensure it is diffused more rapidly, so that any given substitution level is reached sooner. By symmetry, inferior methods for producing the old material have exactly the same effect.

As a second example, consider the impact of a greater willingness on the part of users to switch to the new commodity, as reflected in a higher value for β. Such a change has no effect on the equilibrium niches, nor the values of RA and s_1. But it does increase BC_n and hence the rapidity with which diffusion and substitution take place. Exactly similar implications follow from an increase in the accumulation coefficient f.

Using this method, all manner of factors affecting the substitution process may be made explicit and precise in terms of a comparison between different substitution curves. Moreover, such exercises provide a basis not simply for comparing different substitution processes but for discussing the more complex consequences of changes in parameters during the substitution process. Taking account of such changes, we can generate a family of logistic substitution curves which typically form a non-logistic substitution envelope. The precise shape of this envelope depends on the nature and magnitude of the parameter changes and their temporal incidence. The cautionary implication is clear. Without a knowledge of the changes in determining conditions during the diffusion process, it is not possible to infer the dynamics of substitution from an empirical knowledge of the substitution envelope. Even simple logistic substitution, processes of the kind discussed above, may be associated with markedly non-logistic diffusion and substitution envelopes.

The framework of our Marshallian model is perhaps too simplistic. We have not allowed for foreign trade (Metcalfe and Soete, 1984) or for situations in which the old technology is not in its long-run niche when the innovation occurs (Metcalfe and Gibbons, 1987). Nonetheless, this simple model does make explicit the factors shaping the diffusion and substitution process within the competitive context.

Conclusions

In this chapter we have outlined alternative approaches to the study of innovation diffusion, and suggested that diffusion phenomena should be treated as part of the broader picture of competition and structural change in capitalist economies. At the root of our review is the notion that competition is driven by technological variety, variety which is evaluated in economic terms by the prevailing market environment. Both equilibrium and disequilibrium methods for analysing diffusion play a role in defining niches for technologies, and the paths of adjustment towards these niches.

We have not treated adequately the complex question of the link between the diffusion process and the mechanisms which generate technological variety. Possibly the ideas presented here will provide one foundation for such a study but this task remains unaccomplished.

References

Barzel, Y. (1968), 'The optimum timing of innovations', *Review of Economics and Statistics*, vol. 50.

Binswanger, H. and Ruttan, V. (1978), *Induced Innovation*, Baltimore, Johns Hopkins University Press.

Burns, A. F. (1934), *Production Trends in the United States Since 1870*, NBER.

Burns, T. and Stalker, G. (1961), *The Management of Innovation*, London, Tavistock.

Cameron, H. and Metcalfe, J. S. (1987), 'On the economics of technological substitution', *Technological Forecasting and Social Change*, vol. 31.

Carter, C. and Williams, B. (1957), *Industry and Technical Progress*, Oxford University Press, Oxford.

Czepiel, Z. (1975), 'Patterns of interorganizational communication and the diffusion of a major technological innovation in a competitive industrial community', *Academy of Management Journal*, vol. 18.

David, P. (1975), 'The mechanization of reaping in the anti-bellum Midwest', in *Technical Innovation and Economic Growth*, Cambridge, Cambridge University Press.

David, P. and Olsen, T. (1984), 'Anticipated automation: a rational expectations model of technology diffusion', Stanford, mimeo.

Davies, S. (1979), *The Diffusion of Process Innovations*, Cambridge, Cambridge University Press.

Dosi, G. (1986), 'Institutions and markets in a dynamic world', SPRU, University of Sussex, mimeo.

Downie, J. (1955), *The Competitive Process*, London, Duckworth.

Farrell, M. (1957), 'The measurement of productive efficiency', *Journal of Royal Statistical Society*, (A), vol. 120.

Frankel, M. (1955), 'Obsolescence and technological change', *American Economic Review*, vol. 45.

Feller, I. (1966), 'The draper loom in New England Textiles, 1894–1914: a study of diffusion of an innovation', *Journal of Economic History*, vol. 26.

Georghiou, L. *et al.* (1986), *Post Innovative Performance*, London, Macmillan.

Gold, B. (1981), 'Technological diffusion in industry: research needs and shortcomings', *Journal of Industrial Economics*, vol. 30.

Graham, G. S. (1956), 'The ascendancy of the sailing ship', *Economic History Review*, vol. 9.

Griliches, Z. (1957), 'Hybrid corn: an exploration in the economics of technological change', *Econometrica*, vol. 25.

Hannan, M. and Freeman, J. (1977), 'The population ecology of organizations', *American Journal of Sociology*, vol. 82.

Harley, C. K. (1975), 'On the persistence of old technologies', *Journal of Economic History*, vol. 33.

Hunter, L. (1949), *Steamboats on the Western Rivers*, Cambridge, Harvard University Press.

Iwai, K. (1984), 'Schumpeterian dynamics, Parts I and II', *Journal of Economic Behaviour and Organization*, vol. 6.

Kuznets, S. (1929), *Secular Movements in Prices and Production*, Boston, Houghton Mifflin.

Lawrence, P. and Lorsch, J. (1967), *Organization and Environment*, Cambridge, Mass., Harvard University Press.

Layton, E. (1974), 'Technology as knowledge', *Technology and Culture*, vol. 15.

Lazonick, W. (1983), 'Industrial organization and technological change: the decline of the British cotton industry', *Business History Review*, vol. 57.

Linstone, D. and Sahal, D. (1976), *Technological Substitution: Forecasting Technologies and Applications*, New York, Elsevier.

Lutz, F. and Lutz, V. (1951), *The Theory of Investment of the Firm*, Princeton, Princeton University Press.

Mahajan, V. and Peterson, R. A. (1985), *'Models for Innovation Diffusion'*, New York, Sage.

Mansfield, E. (1961), 'Technical change and the rate of imitation, *Econometrica*, vol. 29.

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

— (1984), 'Technological innovation and the competitive process', in P. Hall (ed.), *Technology, Innovation and Economic Policy,* London, P. Allen.

Metcalfe, J. S. and Gibbons, M. (1987), 'On the economics of structural change and the evolution of technology', in L. Pasinetti (ed.), *Growth and Structural Change*, London, Macmillan.

Metcalfe, J. S. and Soete, L. (1984), 'Notes on the evolution of technology and international competition', in M. Gibbons *et al.* (eds), *Science and Technology Policy in the 1980s and Beyond*, London, Longman.

Monod, J. (1983), *Chance and Necessity,* Hammondsworth, Penguin Books.

Nelson, R. (1968), 'A diffusion model of international productivity differences in manufacturing industry', *American Economic Review*, vol. 58.

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

Pavitt, K. (1983), 'Patterns of technical change: evidence, theory and policy implications', *Papers in Science Technology and Public Policy*, SPRU.

Price, G. R. (1972), 'Extension of covariance selection mathematics', *Annals of Human Genetics*, vol. 35.

Oster, S. (1982), 'The diffusion of innovation among steel firms: the basic oxygen process', *Bell Journal of Economics*, vol. 13.

Ray, G. (1984), *The Diffusion of Mature Technologies*, Cambridge, Cambridge University Press.

Richardson, G. B. (1972), 'The organization of industry', *Economic Journal*, vol. 82.

Rogers, E. M. (1983), *Diffusion of Innovations*, 3rd edn., New York, Free Press.

Rosenberg, N. (1976), 'On technological expectations', *Economic Journal*, vol. 86.

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

Rosseger, G. (1979), 'Diffusion and technological specificity', *Journal of Industrial Economics*, vol. 28.

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

Sandberg, L. (1984), 'The entrepreneur and technological change', in R. Floud and D. McCloskey (eds), *The Economic History of Britain Since 1700*, vol. 2, Cambridge, Cambridge University Press.

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

Schuster, P. and Sigmund, K. (1983), 'Replicator dynamics', *Journal of Theoretical Biology*, vol. 100.

Silverberg, G. (1987), 'Technical progress, capital accumulation and effective demand: a self-organizational model', in D. Batten, J. Costi and B. Johansson (eds), *Economic Evolution and Structural Adjustment,* Berlin, Heidelberg, New York, Tokyo, Springer-Verlag.

Soete, L. and Turner, R. (1984), 'Technology diffusion and the rate of technological change', *Economic Journal*, vol. 94.

Steindl, J. (1952), *Maturity and Stagnation in American Capitalism*, Monthly Review Press. New York.

Stinchcombe, L. (1965), 'Social structure and organizations', in J. G. March (ed.), *Handbook of Organizations*, New York, Rand McNally.

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

Stoneman, P. and Ireland, N. (1983), 'The role of supply factors in the diffusion of new process technology', *Economic Journal*, vol. 93.

Thirtle, C. and Ruttan, V. (1987), *The Role of Demand and Supply in the Generation and Diffusion of Technological Change*, vol. 21, Fundamentals of Pure and Applied Economics, New York, Harwood.

Vincenti, W. de (1985), 'Technological knowledge without science: the innovation of flush riveting in American airplanes', *Technology and Culture*, vol. 25.

26 Competing technologies: an overview*

W. Brian Arthur

Food Research Institute, Stanford University, Stanford, California

> Every steam carriage which passes along the street justifies the confidence placed in it; and unless the objectionable feature of the petrol carriage can be removed, it is bound to be driven from the road, to give place to its less objectionable rival, the steam-driven vehicle of the day.
>
> William Fletcher (1904), *Steam Carriages and Traction Engines*, p. xi.

Introduction

When a new engineering or economic possibility comes along, usually there are several ways to carry it through. In the 1890s the motor carriage could be powered by steam, or by gasoline, or by electric batteries. In more modern times nuclear power can be generated by light-water, or gas-cooled, or heavy-water, or sodium-cooled reactors. Solar energy can be generated by crystalline-silicon or amorphous-silicon technologies. An AIDS vaccine may eventually become possible by cell-type modification methods, or by chemical synthesis, or by anti-idiotype methods. Video recording can be carried out by Sony Betamax® or by VHS technologies.

In each case we can think of these methods or technologies as 'competing' for a 'market' of adopters (Arthur, 1983). They may 'compete' unconsciously and *passively*, like species compete biologically, if adoptions of one technology displace or preclude adoptions of its rivals. Or they may compete consciously and *strategically*, if they are products that can be priced and manipulated. (In this latter case, following nomenclature introduced in Arthur (1985), we will say they are *sponsored*).

What makes competition between technologies interesting is that usually technologies become more attractive—more developed, more widespread, more useful—the more they are adopted. Thus competition between technologies usually becomes competition between bandwagons, and adoption markets display both a corresponding instability and a high degree of unpredictability.

Increased attractiveness caused by adoption, or what I will call 'increasing returns to adoption', can arise from several sources; but five are particularly important:

---

* I am grateful to Paul David, Giovanni Dosi, Frank Englmann, Christopher Freeman, Richard Nelson, Nathan Rosenberg, Gerald Silverberg and Luc Soete for comments on this chapter, and to participants at the May 1987 IFIAS meeting on Technical Change and Economic Theory, Maastricht, The Netherlands.

(i) *Learning by using* (Rosenberg, 1982). Often the more a technology is adopted, the more it is used and the more is learned about it; therefore the more it is developed and improved. A new airliner design, like the DC-8, for example, gains considerably in payload, passenger capacity, engine efficiency and aerodynamics, as it achieves actual airline adoption and use.

(ii) *Network externalities* (Katz and Shapiro, 1985). Often a technology offers advantages to 'going along' with other adopters of it—to belonging to a network of users. The video technology VHS is an example. The more other users there are, the more likely it is that the VHS adopter benefits from a greater availability and variety of VHS-recorded products.

(iii) *Scale economies in production*. Often, where a technology is embodied in a product, like the polaroid technology, the cost of the product falls as increased numbers of units of it are produced. Thus the technology can become more attractive in price as adoption increases.

(iv) *Informational increasing returns*. Often a technology that is more adopted enjoys the advantage of being better known and better understood. For the risk-averse, adopting it becomes more attractive if it is more widespread.

(v) *Technological interrelatedness* (Frankel, 1955). Often, as a technology becomes more adopted, a number of other sub-technologies and products become part of its infrastructure. For example, the gasoline technology has a huge infrastructure of refineries, filling stations, and auto parts that rely on it. This puts it at an advantage in the sense that other technologies, if less adopted, may lack the requisite infrastructure or may require a partial dismantling of the more widespread technology's in-place infrastructure.

Of course, with any particular technology, several of these benefits to increased adoption may be mixed in and present together. Rarely do we have a pure source of increasing returns to adoption.

Whatever the source, if increasing returns to adoption are indeed present, they determine the character of competition between technologies. If one technology gets ahead by good fortune, it gains an advantage. It can then attract further adopters who might otherwise have gone along with one of its rivals, with the result that the adoption market may 'tip' in its favor and may end up dominated by it (Arthur, 1983). Given other circumstances, of course, a different technology might have been favored early on, and *it* might have come to dominate the market. Thus in competitions between technologies with increasing returns, ordinarily there is more than one possible outcome. In economic terms there are multiple equilibria. To ascertain how the *actual* outcome is 'selected' from these multiple candidate outcomes, we need to keep track of how adoptions of rival technologies build up (together with the small events that might influence these) and how they eventually sway and tip the market. We need, in other words, to follow the dynamics of adoption.

Where competing technologies possess increasing returns, a number of very natural questions arise:

1. How can we model the adoption process when there is competition between increasing-return technologies and hence indeterminacy in the outcome?
2. What analytical techniques can be brought to bear on this increasing-return allocation problem? In particular, what techniques can help us determine the possible outcomes of the adoption process?
3. When technologies compete, under what circumstances *must* one technology—albeit an indeterminate one at the outset—achieve a monopoly and eventually take 100 per cent of the adoption market? Under what circumstances will the market eventually be shared?
4. How does the 'competing standards' case differ from the competing technologies one?
5. What difference does it make to have different sources of increasing returns: network externalities rather than learning effects, for example?
6. What policy issues arise in the competing technology case?
7. What major research questions remain to be answered?

In this chapter I have been asked to provide an overview of my work on the competing technology problem, highlighting in particular the dynamic approach. Where possible I will connect my approach and results with those of others and I will mention open research problems. I begin with a review of the basic competing technologies model and then go on to discuss some of the questions raised above.

Lock-in by small events: a review of the basic model

As one possible, simple model of competition between technologies with increasing returns (Arthur, 1983), imagine two unsponsored technologies, *A* and *B*, competing passively for a market of potential adopters who are replacing an old, inferior technology. As adoptions of *A* (or *B*) increase, learning-by-using takes place and improved versions of *A* (or *B*) become available, with correspondingly higher payoffs or returns to those adopting them. Each agent—each potential adopter—must choose either *A* or *B* when his time comes to replace the old technology. Once an agent chooses he sticks to his choice. The versions of *A* or *B* are fixed when adopted, so that agents are not affected by the choices of future adopters.

Suppose for a moment, in a preliminary version of this model, all agents are alike. And suppose that returns to adopting *A* or *B* rise with prior adoptions as in Table 26.1. The dynamics of this preliminary model are trivial but instructive. The first agent chooses the higher payoff technology —*A* in this table. This bids the payoff of *A* upward, so that the next agent *a fortiori* chooses *A*. *A* continues to be chosen, with the result that the adoption process is locked in to *A* from the start. Notice that *B* cannot get a footing, even though if adopted it would eventually prove superior.

Table 26.1 Returns to adopting *A* or *B*, given previous adoptions

Previous adoptions	0	10	20	30	40	50	60	70	80	90	100
Technology *A*	10	11	12	13	14	15	16	17	18	19	20
Technology *B*	4	7	10	13	16	19	22	25	28	31	34

Already in this simple preliminary model, we see two properties that constantly recur with competing technologies: *potential inefficiency* in the sense that the technology that 'takes the market' need not be the one with the longer-term higher payoff to adopters; and *inflexibility*, or lock-in, in the sense that the left-behind technology would need to bridge a widening gap if it is to be chosen by adopters at all.

Although there are examples of technologies that lock out all rivals from the start, this preliminary model is still not very satisfactory. The outcome is either predetermined by whichever technology is initially superior or, if both are evenly matched the outcome is razor-edged. In reality, adopters are not all alike and, at the outset of most competitions, some would naturally prefer technology *A*, and some technology *B*. If this were the case, the order in which early adopter types arrived would then become crucial, for it would decide how the market might 'tip'.

Consider now a full model that shows this. We now allow two types of adopters, *R* and *S*, with 'natural' preferences for *A* and *B* respectively, and with payoffs as in Table 26.2. Suppose each potential-adopter type is equally prevalent, but that the actual 'arrivals' of *R* and *S* agents are subject to 'small unknown events' outside the model, so to speak. Then all we can say is that it is equally likely that an *R* or an *S* will arrive next to make their choice. Initially at least, if an *R*-agent arrives at the 'adoption window' to make his choice, he will adopt *A*; if an *S*-agent arrives, he will adopt *B*. Thus the difference in adoptions between *A* and *B* moves up or down by one unit depending on whether the next adopter is an *R* or an *S*, that is, it moves up or down with probability one-half. This process is a simple gambler's coin-toss random walk. There is only one complication. If by 'chance' a large number of *R*-types cumulates in the line of choosers, *A* will then be heavily adopted and hence improved in payoff. In fact, if *A* gains a sufficient lead over *B* in adoptions, it will pay *S*-types choosing to switch to *A*. Then both *R*- and *S*-types will be adopting *A*, and only *A*, from then on. The adoption process will then become locked in to technology *A*. Similarly, if a sufficient number of *S*-types had by 'chance' arrived to adopt *B*

Table 26.2 Returns to adopting A or B, given n_A and n_B previous adopters of A and B^a

	Technology A	Technology B
R-agent	$a_R + rn_A$	$b_R + rn_B$
S-agent	$a_S + sn_A$	$b_S + sn_B$

[a] The model assumes that $a_R > b_R$ and that $b_S > a_S$. Both r and s are positive.

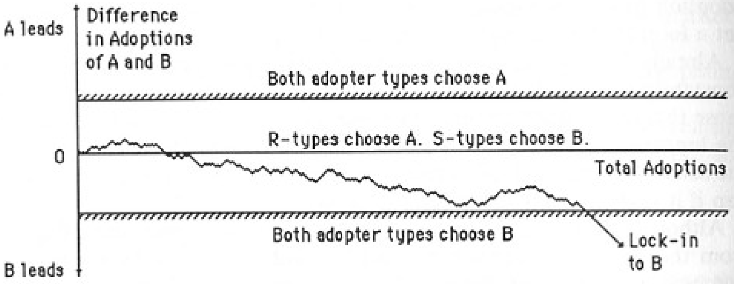


Figure 26.1 Difference in adoptions: random walk with absorbing barriers

over A, B would improve sufficiently to cause R-types to switch to B. The process would instead lock in to B (see Figure 26.1). Our random walk is really a random walk with absorbing barriers on each side, the barriers corresponding to the lead in adoption it takes for each agent-type to switch its choice.

All this is fine. We can now use the well-worked-out theory of random walks to find out what happens to the adoption process in the long run. The important fact about a random walk with absorbing barriers is that absorption occurs eventually with certainty. Thus in the model I have described, the economy *must* lock in to monopoly of one of the two technologies, A or B, but *which* technology is not predictable in advance. Also, the order of choice of agent is not 'averaged away'. On the contrary, it decides the eventual market outcome. Thus, the process is *non-ergodic*—or more informally we can say that it is *path-dependent* in the sense that the outcome depends on the way in which adoptions build up, that is, on the path the process takes. As before, the process becomes inflexible: once lock in occurs the dominant technology continues to be chosen; hence it continues to improve, so that an ever larger boost to the payoff of the excluded technology would be needed to resuscitate it. Further, it is easy to construct examples in which this 'greedy algorithm' of each agent taking the technology that pays off best at his time of choice may miss higher rewards to the future adoption and development of the excluded technology. As in the preliminary model, economic efficiency is not guaranteed.

This model, like all theoretical models, is obviously stylized. But it does capture an important general characteristic of competition between technologies with increasing returns. Where the competition is not dead at the outset, with a single technology dominating from the start, the adoption process is inherently unstable, and it can be swayed by the cumulation of small 'historical' events, or small heterogeneities, or small differences in timing. Thus low-level events, stemming from the inevitable graininess present in the economy, can act to drive the process into the 'gravitational orbit' of one of the two (or, with several technologies competing, many) possible outcomes. What we have in this simple model is 'order' (the eventual adoption-share outcome) emerging from 'fluctuation' (the inherent randomness in the arrival sequence). In modern terminology, our competing-technologies adoption process is therefore a *self-organising process* (Prigogine, 1976). (For further details on self-organisation, see the chapters of Silverberg, Corncelli and Dosi, and Dosi and Orsenigo in this volume).

Of course, it could be objected that at some level—in some all-knowing Laplacian world—the arrival sequence in our model is fore-ordained, and that therefore the outcome that this sequence implies is fore-ordained, and that therefore our technology competition is determinate and predictable. Ultimately this comes down to a question of modelling strategy. Where increasing returns are present, different patterns of small events—whether known or not—can lead to very different outcomes. If they are unknown at the outset, if for practical purposes they lie beneath the resolution of our model, we must treat them as random; so that unless we believe we know all events that can affect the build-up of adoptions and can therefore include them explicitly, models of technological competition must typically include a random component. In the model above, randomness was introduced by lack of knowledge of the arrival sequence of the adopters. But in other models it could have different sources. Randomness might, for example, enter in a homogeneous adopter-type model because technological improvements occur in part by unpredictable breakthroughs. The subject is new enough that even obvious extensions like this have not yet been studied. There may be a wide class of competing-technology models, but we would expect to see much the same properties as we found above upheld: inflexibility or lock-in of outcome; non-predictability; possible inefficiency; and non-ergodicity or path-dependence.

Do real-world competitions between technologies show these properties? Does the economy sometimes lock in to an inferior technology because of small, historical events? It appears that it does. Light-water reactors at present account for close to 100 per cent of all US nuclear power installations and about 80 per cent of the world market. They were originally adapted from a highly compact unit designed to propel the first American nuclear submarine, the USS *Nautilus*, launched in 1954 (Weinberg, 1954).

A series of circumstances—among them the US Navy's role in early construction contracts, political expediency within the National Security Council, the behavior of key personages like Admiral Rickover, and the Euratom Program—acted to favor light water, so that learning and construction experience gained with light water early on locked the market in by the mid-1960s (Cowan, 1987). And yet the engineering literature consistently argues that, given equal development, the gas-cooled design would have been superior (Agnew, 1981).

Similarly, gasoline now dominates as the power source for automobiles. It may well be the superior alternative, but certainly in 1895 it was held to be the least promising option. It was hard to obtain in the right grade; it was dangerous; and it required more numerous and more sophisticated moving parts than steam. Throughout the period 1890–1920, developers, with predilections depending on their previous engineering experience, produced constantly improving versions of the steam, gasoline and electric automobiles. But a series of circumstances—among them, in the North American case, unlikely ones like a 1895 horseless carriage competition which appears to have influenced Ransom Olds in his decision to switch from steam to gasoline, and an outbreak in 1914 of hoof-and-mouth disease that shut down horse troughs where steam cars drew water (McLaughlin, 1954; Arthur, 1984)—gave gasoline enough of a lead that it subsequently proved unassailable. Whether steam and electric cars, given equal development, could have been superior is not clear; but this question remains under constant debate in the engineering literature (Burton, 1976; Strack, 1970).

Is lock-in to a possibly inferior technology permanent? Theoretically it is, where the source of increasing returns is learning-by-using, at least until yet newer technologies come along to render the dominant one obsolete. But lock-in need not be permanent if network externalities are the source. Here, if a technology's advantage is mainly that most adopters are 'going along' with it, a coordinated changeover to a superior collective choice can provide escape. In an important paper, Farrell and Saloner (1985) showed that as long as agents know other agents' preferences, each will decide independently to 'switch' if a superior alternative is available. But where they are uncertain of others' preferences and intentions, there can be 'excess inertia': each agent would benefit from holding the other technology but individually none dares change in case others do not follow.

Whatever the source of increasing returns in competitions between technologies, the presence of lock-in and sudden release causes the economy to lose a certain smoothness of motion.

Technology structure: the path-dependent Strong Law of Large Numbers

In the discussion so far, we have derived some basic ideas and properties of technology competition from a dynamic model with a very particular

linear-returns-from-learning mechanism. We would like to be able to handle competing-technology problems with more general assumptions and returns-to-adoption mechanisms. In particular we are interested in qualitative questions such as whether, and under what circumstances, an adoption market must end up dominated by a single technology.

In thinking about the type of analytical framework we would need for more general versions of the problem, it seems important to preserve two properties: (i) that choices between alternative technologies are affected by the numbers of each alternative present in the adoption market at the time of choice; equivalently, that choices are affected by current market shares; (ii) that small events outside the model may influence the process, so that a certain amount of randomness must be allowed for. Thus the 'state' of the market may not determine the next choice, but rather the probability of each alternative being chosen.

Consider a dynamical system that abstracts and allows for these two properties. I will call it an *allocation process*. At each time that a choice occurs, a unit addition or allocation is made to one of K categories, with probabilities $p_1(x)$, $p_2(x)$ . . ., $p_K(x)$, respectively, where this vector of probabilities p is a function of x, the vector giving the proportion of units currently in categories 1 to K (out of the total number n so far in all categories). In our competing technologies problem, this corresponds to a choice of one technology from K competing alternatives, each 'time' of choice, with probabilities that depend upon the numbers of each alternative already adopted and therefore upon current adoption shares.[1] (For a given problem, if we know the source of randomness and the payoff-returns at each state of the market, we can, in principle at least, derive these probabilities as a function of adoption shares.)

Our question is: what happens to the long-run proportions (or adoption shares) in such a dynamical system? What long-run technological structures can emerge? The standard probability-theory tool for this type of problem is the Strong Law of Large Numbers which makes statements about long-run proportions in processes where increments are added at successive times. For example, if we successively add a unit to the 'category' Heads with probability 1/2 in tossing a coin, the standard Strong Law tells us that the proportion of Heads must settle to 0.5. But we cannot use the standard Strong Law in our process. We do not have the required *independent* increments. Instead we have increments—unit adoptions or allocations to technologies 1 through K—which occur with probabilities influenced by past increments. We have a 'coin' whose probability of Heads changes with the proportion of Heads tossed previously.

We can still generate a Strong Law for our dependent-increment process. Suppose we consider the mapping from present proportions, or adoption shares, to the probability of adoption, as with the two examples in Figure 26.2, where $K = 2$. We can see that where the probability of adoption A is higher than its market share, there would be a tendency in the allocation (or adoption) process for A to increase in proportion; and

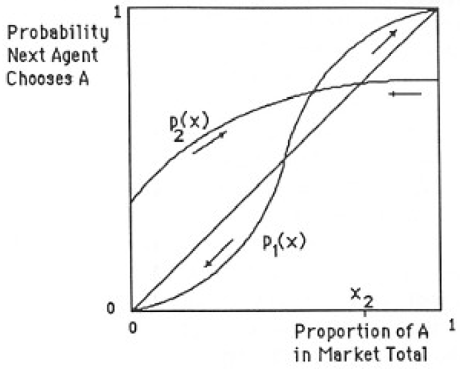


Figure 26.2 Probability of adoption as a function of adoption share

where it is lower, there would be a tendency for it to decrease. If the proportions or shares in each category settle down as total allocations increase, then they should settle down at a fixed point of this mapping. In 1983 Arthur, Ermoliev and Kaniovski proved that (under certain technical conditions) indeed this conjecture is true.^{2} Allocation processes indeed settle down in the long run, with probability one, to an unchanging vector of proportions (adoption shares) represented by one of the fixed points of the mapping from proportions (or adoption shares) to the probability of adoption. They converge to a vector of adoption shares x where $x = p(x)$. Not all fixed points are eligible. Only 'attracting' or stable fixed points (ones that expected motions lead toward) can emerge as the long-run outcomes. (Thus in Figure 26.2 the possible long-run shares are 0 and 1 for the function p_1 and x_2 for the function p_2.) Of course, where there are multiple fixed points, *which* one is chosen depends on the path taken by the process: it depends upon the cumulation of the random events that occur along the way. This very general Strong Law for dependent-increment processes (which, following convention, I shall label 'AEK') generalizes the conventional Strong Law of Large Numbers.

The *allocation process* framework, with its corresponding Strong Law, applies to a wide variety of self-organizing or autocatalytic problems in economics and physics (Arthur, Ermoliev and Kaniovski, 1984, 1987; Arthur, 1986, 1987). For our competing-technology purposes, however, we now have a powerful piece of machinery that enables us to investigate the possible long-run adoption outcomes under different adoption-market mechanisms. For a particular problem we would proceed in three steps:

1. Detail the particular mechanisms at work in the adoption process, paying special attention to returns functions, heterogeneities, and sources of randomness.

2. Use this knowledge to derive the probabilities of choice of each technology explicitly as a function of current adoption shares.
3. Use the AEK Strong Law to derive actual long-run possible adoption shares as the stable fixed points of the adoption-share-to-probability mapping.

A number of studies now use this technique (see, for example, David, 1986). The 'informational increasing returns' model of Arthur (1985b) is an example. In this model risk-adverse potential adopters are uncertain about the actual payoff of two fixed payoff technologies they can choose from. They gather information by 'polling' some random sample of previous adopters. (Neither learning-by-using nor network effects are present.) Increasing returns come about because, if adopters of A are more numerous, the next chooser will likely sample more A's than B's, and will therefore be better informed on A. Being risk-averse, he will therefore choose A with a probability that increases with A's proportion of x of the market. Application of the AEK Strong Law to a rigorous model of this mechanism yields precise circumstances under which 'informational increasing returns' allow stable fixed points only at the points $x = 0$ and $x = 1$. That is, it yields circumstances under which informational increasing returns alone cause eventual monopoly of A or of B with probability 1.

When is technological monopoly inevitable?

Is it inevitable that one technology must eventually shut out the others when there are increasing returns to adoption? The answer is no. Consider a more general version of the heterogeneous-adopter-unknown-arrival-sequence model, in which there is now a continuum of agent types rather than just two. We can now think of agents—potential adopters—as distributed over adoption payoffs as in Figure 26.3. An adopter is chosen at random from this probability distribution each time a choice is to be made; and the distribution itself shifts either to the right or upward as returns to A or B increase with an adoption of either A or B respectively. Monopoly—lock-in to a single technology—corresponds to the distribution of payoffs getting 'driven' over the 45° line in this two-technology case. (We assume the distribution of adopter payoffs has 'bounded support'—that is, it does not tail off to infinity in any direction.) Where K technologies compete, we can use the AEK Strong Law to show that where there is no ceiling to the increasing returns (so that returns increase without bound as adoptions increase) then sooner or later one technology *must* by the cumulation of chance achieve sufficient adoption advantage to drive the distribution of adopters 'over the line'. With unbounded increasing returns eventual monopoly by a single technology is indeed inevitable (Arthur, 1986).

But where returns to adoption increase but are bounded, as when learning effects eventually become exhausted, monopoly is no longer

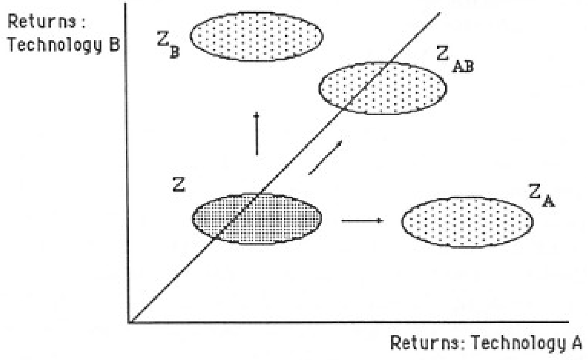


Figure 26.3 Payoffs to adoption of A and B under a continuum of adopter types[a]

[a] At the outset adopter payoffs lie in set Z. Adoptions of technology A only shift this set horizontally to the right as in Z_A. Adoptions of B only shift it vertically as in Z_B. Adoptions of A and B shift it diagonally as in Z_{AB}.

inevitable. The reason is interesting. In this case, certain sequences of adopter types could bid the returns to both technologies upward more or less in concert. These technologies could then reach their 'increasing returns ceilings' together, with adopter-type-payoffs still straddled across the 45° line (as with Z_{AB} in Figure 26.3), and thus with the adoption market shared from then on. But other adopter-arrival sequences may push the payoff distribution across the line early on. Thus with increasing returns to adoption that are bounded, the general finding is that some 'event histories' dynamically will lead to a shared market; other event histories lead to monopoly (Arthur, 1986).

Exact conditions for monopoly in the strategic-competition case where technologies exist as 'sponsored' products are not yet known. Hanson (1985) explored a version of this IBM-versus-Apple problem, building on the basic linear-increasing-returns model above. He assumed that firms could price technologies and thereby manipulate adoption payoffs in a market where heterogeneous adopters arrived at random. Hanson was able to show in this stochastic-duopoly problem that firms would price low early on to gain adoptions, possibly even taking losses in an arm-wrestling match for market share. If both firms were evenly matched enough to stay in the market under these circumstances, then sooner or later the cumulation of 'chance events' might allow one firm sufficient adoption advantage to tip the market in its favor. It would then have sufficient advantage to be able to raise its price and take monopoly profits, while keeping the other firm on the contestable margin of the market. Using AEK, Hanson was able to detail certain conditions under which monopoly by a single-technology-product would be inevitable. It is clear, however, that conditions can be constructed where the markets can also end up shared. For example, when increasing returns are bounded, and firms discount future income heavily so that they are mainly interested in present sales, neither firm may wish to price low early on. Neither might then eventually win the 'natural customers' of the other and the result would be a shared market.

Competing standards and the role of expectations

The term 'standard' has two meanings in the technology literature: that of a convention or code of practice, such as distributing alternating current at 110 volts or transmitting it at 60 hertz; and that of the technology or method or code that comes to dominate—that becomes 'standard'. Standards in the first sense—conventions—can compete much the same as method-technologies do, for a market of adherents, or users, or adopters. Competing standards raise somewhat different issues from competing technologies (see David, 1986, and the papers of Katz and Shapiro, and Farrell and Saloner). I will treat standards here only in so far as they overlap with our dynamics-of-adoption problem.

With standards, learning, information and production externalities are less important and the main sources of increasing returns are network externalities and possibly technological interrelatedness. Both sources confer benefits if *future* adopters go along with one's choice. This introduces something not yet considered in our discussion—*expectations*.

Katz and Shapiro, in an important paper (1985), consider a static version of the problem of competing 'networks' of different standards, in which 'network externalities' accrue to increased network size. The networks are provided by firms which must determine network size in advance. It pays firms to provide large networks if potential adopters expect these networks to be large and thereby commit their choice to them. Therefore if, prior to adoption, sufficient numbers of agents believe that network A will have a large share of adopters, it will; but if sufficient believe B will have a large share, *it* will. Katz and Shapiro showed that there could be multiple 'fulfilled-expectation equilibria' that is, multiple sets of eventual network adoption shares that fulfil prior expectations.

In this simple but important model, expectations are given and fixed before the adoption process takes place. More realistically, if adoption were not instantaneous, potential adopters might change or modify their expectations as the fortunes of alternatives changed during the adoption process itself. One possible formulation (Arthur, 1985a) is to assume that agents form expectations in the shape of beliefs about the adoption process they are in. That is, they form probabilities on the future states of the adoption process—probabilities that are conditioned on the numbers of current adoptions of the competing alternatives. Thus these probabilities, or beliefs, change and respond as the adoption market changes. (We would have a *fulfilled-equilibrium-stochastic-process* if the *actual* adoption process that results from agents acting on these beliefs turns out to have conditional probabilities that are identical to the *believed* process.) In this model, if one standard, or technology, gets ahead by 'chance' adoptions, its increased probability of doing well in the adoption market will further enhance expectations of its success. Analysis here confirms the basic Katz and Shapiro finding. Adaptive or dynamic expectations act to destabilize further an already unstable situation: lock-in to monopoly positions now occurs more easily.

Policy issues

We have seen that in uncontrolled competitions between technologies with learning effects, or network externalities, or other sources of increasing returns to adoption, there is no guarantee that the 'fittest' technology—the one with superior, long-run potential—will survive. There are therefore grounds for intervention.

Where a central authority with full information on future returns to alternative adoption paths knows which technology has superior long-run potential, it can of course attempt to 'tilt' the market in favor of this technology. Timing is, of course, crucial here (Arthur, 1983): in Paul David's phrase (1986) there are only 'narrow windows' in which policy would be effective.

More often, though, it will not be clear in advance which technologies have most potential promise. The authorities then face the difficult problem of choosing which infant technologies to subsidize or bet on. This yields a version of the multi-arm bandit problem (in which a gambler plays several arms of a multi-arm bandit slot machine, trying to ascertain which has the highest probability of producing jackpots). Cowan (1987) has shown that, where central authorities subsidize increasing-return technologies on the basis of their current estimates of future potential, locking into inferior technologies is less likely than in the uncontrolled adoption case. But it is still possible. An early run of bad luck with a potentially superior technology may cause the central authority, perfectly rationally, to abandon it. Even with central control, escape from inferior technological paths is not guaranteed. This finding is important for projects like the US Strategic Defense Initiative, where ground-based excimer lasers, particle-beam weapons, X-ray lasers, homing vehicles and other devices compete for government subsidy on the basis of expected long-run promise. Where each of these improves with development, it is likely that lock-in to one will occur; however it may not be lock-in to the one with superior long-run potential.

It may sometimes be desirable as a policy option to keep more than one technology 'alive', to avoid monopoly problems (if the technology is marketed), or to retain 'requisite variety' as a hedge against shifts in the economic environment or against future 'Chernobyl' revelations that the technology is unsafe. The question of using well-timed subsidies to prevent the adoption process 'tipping' and shutting out technologies has not yet been looked at. But its structure—that of artificially stabilising a naturally unstable dynamical process—is a standard one in stochastic feedback control theory.

Some research questions

Several open or only partially resolved research questions have already been mentioned. Besides these, there are at least three major classes of problems that I believe would benefit from future study:

1. *Recontracting models*. Where the sources of increasing returns is learning-by-using, results would change little if adopters could re-enter the 'queue' and change their choice at a future date. What counts with learning is the previous number of adoptions of a technology, not the fact that an agent is choosing a second time.[3] Where the source is network externalities, results *would* change substantially. In this case, with agents changing their preferences occasionally as well as striving to go along with the more prevalent alternative, recontracting or changing choice would take place as adoptions built up and might continue even when the market was at its full, saturated size. We would then have something akin to a stochastic version of the Farrell and Saloner (1985) model. The important difference from our earlier models is that with 'deaths' as well as 'births' of adoptions allowed, increments to market-share position would tend to be of constant order of magnitude. Adoption processes with recontracting would therefore tend to show convergence in distribution rather than strong convergence, with 'punctuated equilibria' possible in the shape of long sojourns near or at monopoly of one technology coupled with intermittent changeover to monopoly by a different technology. This type of structure has counterparts in genetics, sociology and in far-from-equilibrium thermodynamics (see Haken, 1978; Weidlich and Haag, 1983). But it has not yet been studied in the technology context.
2. *Empirical studies*. So far we have two excellent historical studies on the set of events and varied sources of increasing returns that led to dominance of the QWERTY typewriter keyboard (David, 1985) and the dominance of alternating current (David and Bunn, 1987). For most present-day uses, alternating current indeed appears to be superior to the alternative, direct current. The QWERTY keyboard, however, may be slightly inferior to the alternative Dvorak keyboard. Norman and Rumelhard (1983) find Dvorak faster by 5 per cent. Missing as yet, however, are detailed empirical studies of the actual

choice-by-choice dynamics of technological competitions. For prominent competitions such as that between nuclear reactors it might be possible to put together a complete account of the adoption sequence and the events that accompanied it. This would allow identification and parameter estimation of the stochastic dynamics of an *actual* rather than a theoretical case.

3. *Spatial technological competition*. One of the striking features of the classical technology diffusion literature (Griliches, 1957; David, 1969) is its concern with the spatial dimension—with the fact that a technology diffuses geographically as well as temporally. In the *competing* technologies problem, geographical diffusion would of course also be present. The spatial dimension would become particularly important if returns to adoption were affected by neighbors' choices. This was the case historically in competitions between railroad gauges (Puffert, 1988) where it was advantageous to adopt a gauge that neighboring railroads were using. The dynamics of spatial-technology competitions have not been explored yet. But they would resemble those of the well-known Ising model in physics and voter models in probability theory, where dipoles and voters respectively are influenced by the states of their nearest neighbors (Liggett, 1979). Here geographical clusters of localities locked in to different technologies might emerge, with long-run adoption structure depending crucially on the particular spatial increasing-returns mechanism at work.

Conclusion

In the classical literature on the economics of technology, a new and superior technology competes to replace an old and inferior one. In this new literature, two or more superior technologies compete with *each other*, possibly to replace an outmoded one. Competition assumes a stronger form. In the competing technologies problem, the theory that emerges is a theory of non-convex allocation. There are multiple equilibria—multiple possible long-run adoption-share outcomes. The cumulation of small 'random' events drives the adoption process into the domain of one of these outcomes, not necessarily the most desirable one. And the increasing-returns advantage that accrues to the technology that achieves dominance keeps it locked in to its dominant position.

I have indicated that competing technologies are examples of self-organising, order-through-fluctuation systems. They are also examples of evolutionary systems, although the mechanisms are quite different from the ones in Nelson and Winter (1982). Where competing technologies possess increasing returns to adoption, one technology can exercise 'competitive exclusion' on the others; if it has a large proportion of natural adopters it will have a 'selectional advantage'; and the importance of early events results in a 'founder effect' mechanism akin to that in genetics.

The dynamical picture of the long-term economy that results is less like that of a sphere smoothly rolling on a flat surface, with its point of contact with the ground unique and ever changing, and more like that of a polytope lurching down a slope. Where technologies compete, patterns—technology adoption structures—lock in. But as time passes and new technological competitions come about the old patterns are changed, shaken up and re-formed, and in due course a new one is locked in.

To the extent that this happens, there may be theoretical limits as well as practical ones to the predictability of the economic future.

Notes

1. If these probabilities depend on *numbers adopted* rather than directly on market shares we can write them as $(p_1(nx), p_2(nx), p_K(nx)$. This becomes equivalent to a probability function p_n that depends on 'time' n as well as adoption share x.
2. Hill, Lane and Sudderth proved a version of this theorem in 1980 for the case $K = 2$ and p unchanging with time n. The informally stated version in the text holds for $K \geqslant 2$ and for time-varying functions p_n provided they converge to a limiting function p (Arthur, Ermoliev and Kaniovski, 1983, 1984, 1986). Technically, the sequence of Borel functions p_n needs to converge to p at a rate faster than $1/n$ converges to zero; the set of fixed points of p needs to have a finite number of connected components; and for $K > 2$ convergence to a point rather than to a cycle or more complex attractor requires the deterministic dynamics formed by the expected motion of the process to be a gradient system. The 1986 paper is perhaps the best introduction to this theorem.
3. Agent arrivals, if second-time choosers, might, however, be dependent on the previous arrival sequence

References

Agnew, H.M. (1981), 'Gas-cooled nuclear power reactors', *Scientific American*, vol. 244, pp. 55–63.

Arthur, W.B. (1983), 'Competing technologies and lock-in by historical events: the dynamics of allocation under increasing returns', International Institute for Applied Systems Analysis, Paper WP-83-90, Laxenburg, Austria.

—— (1984), 'Competing technologies and economic prediction', *Options*, International Institute for Applied Systems Analysis, Laxenburg, Austria.

—— (1985a), 'Competing technologies and lock-in by historical events: the dynamics of allocation under increasing returns', revised from 1983 paper as Center for Economic Policy Research, Paper 43, Stanford, *Economic Journal*, 99 March 1989.

—— (1985b), 'Information, imitation and the emergence of technological structures', Stanford, mimeo.

—— (1986), 'Industry location and the importance of history', Center for Economic Policy Research, Paper 84, Stanford.

—— (1987), 'Urban systems and historical path-dependence', Chapter 4 in R.

Herman and J. Ausubel (eds.) *Cities and Their Vital Systems*, National Academy Press, Washington, D.C.

Arthur, W.B., Ermoliev, Yu. M., and Kaniovski, Yu. M. (1983), 'On generalized urn schemes of the polya kind' (in Russian), *Kibernetika*, vol. 19, pp. 49–56; English trans. in *Cybernetics*, vol. 19, pp. 61–71.

—— (1984), 'Strong Laws for a class of path-dependent urn processes', in Arkin, Shiryayev and Wets (eds.), *Proceedings of the International Conference on Stochastic Optimization, Kiev 1984*, Berlin, Heidelberg, New York, Tokyo, Springer.

—— (1987), 'Path-dependent processes and the emergence of macro-structure', *European Journal of Operational Research*, vol. 30, pp. 294–303.

Burton, Rodney L. (1976), 'Recent advanced in vehicular steam engine efficiency', Society of Automotive Engineers, Preprint 760340.

Cowan, R. (1987), 'Backing the wrong horse: sequential technology choice under increasing returns,' Ph.D.diss., Stanford.

David, P. (1985), 'Clio and the economics of QWERTY', *American Economic Review, Proceedings,* vol. 75, pp. 332–7.

—— (1969), 'A contribution to the theory of diffusion', Memo 71, Stanford Center for Economic Growth, Stanford.

—— (1986), 'Some new standards for the economics of standardization in the information age', Paper 79, Center for Economic Policy Research, Stanford.

David, P. and Bunn, J. (1987), 'The battle of the systems and the evolutionary dynamics of network technology rivalries', Stanford, mimeo.

Farrell, J. and Saloner, G. (1985), 'Standardization, compatibility and innovation', *Rand Journal of Economics*, vol. 16, pp. 70–83.

—— (1986), 'Installed base and compatibility', *American Economics Review*, vol. 76, pp. 940–55.

—— (1986), 'Standardization and variety', *Economic Letters*, vol. 20, pp. 71–4.

Fletcher, W. (1904), *English and American Steam Carriages and Traction Engines*, reprinted, Devon, David & Charles, 1973.

Frankel, M. (1955), 'Obsolescence and technological change in a maturing economy', *American Economic Review*, vol. 45, pp. 296–319.

Griliches, Z. (1957), 'Hybrid corn: an exploration in the economics of technological change', *Econometrica*, vol. 25, pp. 501–22.

Hanson, W. (1985), 'Bandwagons and orphans: dynamic pricing of competing systems subject to decreasing costs', Ph.D.Diss., Stanford.

Haken, H. (1978), *Synergetics*, Berlin, Heidelberg, New York, Tokyo, Springer.

Katz, M. (1986), 'The economics of standardization in networks industries', Princeton, mimeo.

Katz. M. and Shapiro, C. (1985), 'Network externalities, competition and compatibility', *American Economic Review*, vol. 75, pp. 424–40.

—— (1986), 'Technology adoption in the presence of network externalities', *Journal of Political Economy*, vol. 94, pp. 822–41.

Liggett, T. (1979), 'Interacting Markov processes', in Lect. Notes in Biomath 38, Berlin, Heidelberg, New York, Tokyo, Springer.

McLaughlin, C. (1954), 'The Stanley steamer: a study in unsuccessful innovation', *Explorations in Entrepreneurial Hist.*, vol. 7, pp. 37–47.

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

Norman, D. and Rumelhart, D. (1983), 'Studies of typing from the LNR Research Group', in W. Cooper (ed.), *Cognitive Aspects of Skilled Typewriting*, Berlin, Heidelberg, New York, Tokyo, Springer.

Prigogine, I. (1976), 'Order through fluctuation: self-organization and social system', in E. Jantsch and C.H. Waddington (eds.), *Evolution and Consciousness*, New York, Addison-Wesley.

Puffert, D. (1988), 'Network externalities and technological preference in the selection of railway gauges', Ph.D.dissertation, Stanford, forthcoming.

Rosenberg, N. (1982), *Inside the Black Box: Technology and Economics*, Cambridge, Cambridge University Press.

Strack, W.C. (1970), 'Condensers and boilers for steam-powered cars', NASA Technical Note, TN D-5813, Washington, DC.

Weidlich, W. and Haag, G. (1983), *Concepts and Models of a Quantitative Sociology*, Berlin, Heidelberg, New York, Tokyo, Springer.

Weinberg, A.M. (1954), 'Power reactors', *Scientific American*, vol. 191, pp. 33–9.

27 Formalizing growth regimes

Robert Boyer
Centre d'Etudes Prospectives d'Economie Mathematique Appliquees a la Planification, Centre National de la Recherche Scientifique, ecole des hautes etudes en Sciences Sociales, Paris

Technological change and macroeconomics

This chapter deals with some methods for embodying technical change in macroeconomic models, however simple. In a sense it is a follow-up to Chapters 2 and 4. Both of them discuss the need for an analysis of the conditions which must hold between the technological regime and institutional structures in order to make sustained and relatively stable growth possible, or, conversely, the possible mismatchings at the root of unstable and/or slow growth.

The 'régulation' approach has already tried to elaborate such models, even if they are not totally suitable for the present purpose. In the very beginning of the regulation approach, two sectoral models were built (Aglietta, 1974; Billaudot, 1976). They show the necessity of a connection between the consumption and production goods sectors if accumulation is to be a permanent and relatively stable process. Along the same lines, subsequent works have investigated the specificities of the long post-1945 boom in France (Bertrand, 1978, 1983). In this context, Fordism results from a specific growth regime in which intensive technological change and new forms of social organization promote a complementarity between mass production and consumption, modernization and capital deepening. Furthermore, three theoretical models, corresponding to extensive accumulation, Taylorism and Fordism, have been investigated (Fagerberg, 1984). Similarly, a simple growth model of Kaldorian spirit has been compared with stylized historical facts (Boyer and Coriat, 1987) and then used to analyse the viability of flexible specialization or, alternatively, flexible automation.

In a sense, the model presented here benefits from all these formalizations, but tries to go a little beyond them. On the one hand, the model is simplified and keeps only the core of the relevant mechanisms. It has been suggested (Bertrand, 1983) that an aggregate model can retain a large number of the properties of sectoral/disaggregated analysis. On the other hand, the model is nevertheless more general since it investigates not only Fordist but also a large number of other regimes.

The argument consists of five steps. First, the general hypotheses about technological change and productivity, income distribution and demand formation are presented and lead to a simplified growth model. Second, it is shown that according to the precise features of the productive systems and economic mechanisms, very different growth or crisis regime may exist. Then the predictions of the model are compared to some stylized

facts concerning capitalist economies over one or two centuries, the test of these hypotheses being left to future econometric work. Finally, this framework helps us analyse the present economic and technological transition: are the present transformations promoting growth or, on the contrary, deepening unemployment and/or instability? Brief concluding remarks summarize the major findings and propose some areas for future research.

Formalizing the sources of productivity gains and their sharing: back to growth theories

Ideally, the model should distinguish at least between two sectors producing consumption and investment goods. Nevertheless, previous attempts by Boyer (1975), Bertrand (1983) and Fagerberg (1984) do suggest that most of the arguments can be captured within a one-sector model. Thus an aggregate approach is adopted here. The starting point will be a generalization of a previous formalization (Boyer and Coriat, 1987). We want to study a larger variety of accumulation regimes, intensive as well as extensive, with or without mass consumption.

The main hypotheses in a nutshell

The economy is assumed to be closed—thus applicable at the world level or to a national economy not open to external trade. The model could easily be extended to an open economy, as has already been done in, for example, Boyer and Petit (1981a, 1984). We basically want to formalize both productivity gains and their division between wages and profits, i.e. the simultaneous dynamics of production and consumption. Our model will consist of seven endogenous variables, entering into seven behavioural or accounting equations.

Productivity trends plays a prominent role in the system. They are related to three factors. The first is the intensity of innovation as measured by R & D expenditures, number of patents, or the orientation of technical progress towards labour-saving equipment (variable *INNO*). This is supposed to represent a Schumpeterian explanation of productivity. The second is capital deepening, expressed by the investment/output ratio (variable I/Q). This effect could be termed Salterian, since this variable captures the renewal of capital in a vintage model. Third is a Kaldor–Verdoorn effect, linking productivity to output growth via dynamic increasing returns to scale (variable $\mathring{Q}$). One might imagine learning-by-doing effects, or long-run properties linking the division of labour, productivity and the size of the market. Hence we propose the following productivity equation:

$$P\dot{R} = a' + b' . \frac{I}{Q} + d' . \mathring{Q} + e' \overline{INNO} \qquad b', d', e' > 0 \tag{1'}$$

Investment reacts to the dynamism of household consumption ($\mathring{C}$), according to a traditional Keynesian accelerator effect. But in order to contrast different accumulation regimes, one must add another determinant usually introduced by classical theory: the profit share (PRO/Q). Contemporary research usually combines these two factors: investment is either limited by demand—here restricted to consumption—or by profitability. Furthermore, another Schumpeterian effect might also be of some interest in dealing with technical innovation: if innovations are available, firms will be induced to invest more in new products (variable $INNO$). This leads to the formulation:

$$\frac{I}{Q}=f'+v'.\mathring{C}+u'.\left[\frac{PRO}{Q}\right]+e''.\overline{INNO} \quad v',u',e''>0 \qquad (2')$$

Households' consumption is modelled in a very traditional way. The marginal propensity to consume is supposed to be different for wages and profits (c_1 and c_2 respectively). Since the other mechanisms are related to medium-term trends, no lags are introduced in the equation. Hence we have the following equation:

$$\mathring{C}=c_1.(N\overset{\circ}{.}RW)+c_2.(\overline{Q-\mathring{N}.\mathrm{RW}})+g \quad c_1>c_2>0 \qquad (3')$$

Wage formation has to be richer in order to deal with at least two polar cases: purely competitive determination of real wages and productivity increases shared with wage earners more or less according to what has been called the Fordist capital–labour compromise. The first mechanism is captured by a linear elasticity of real wages with respect to employment variations (parameter l). The logic of a Phillips curve is therefore extended to real wages analysed in the medium term. The second one is described by a second elasticity with respect to productivity trends (parameter k). As usual, a constant term is added to capture any other factor. This leads to the following equation:

$$R\mathring{W}=k'.P\mathring{R}+l'.(\mathring{N}-\mathring{\overline{LF}})+h \quad k\geq 0, l\geq 0 \qquad (4')$$

$\mathring{\overline{LF}}$: exogenous evolution of total labour force

Three accounting identities close the model. The first one describes the national accounts identity about resources and uses of total production. The only difficulty is in converting levels into rates of change, α being the share of consumption in total net output for the previous period. The second relation defines changes in employment as the difference between the rate of output growth and productivity increase. Finally, the last equation says that net output is equal to the sum of profit and wages. Thus we have:

$$\mathring{Q}=\alpha.\mathring{C}+(1-\alpha).\mathring{I} \quad 0\leq\alpha\leq 1 \qquad (5')$$

$$\mathring{N}\approx\mathring{Q}-P\mathring{R} \qquad (6')$$

$$\frac{PRO}{Q}=1-\frac{RW}{PR} \qquad (7')$$

The basic economic ideas are fairly simple, but the analytics are a little bit more complex, as is usual when variables expressed in levels and rates of change are mixed. So the next step is to build an approximate version which can be completely solved and discussed mathematically without the need for simulations or the use of general fixed point theorems.

A simplified model

In order to do this, the first three equations are linearized and written in a less satisfactory form, but one which is easier to solve. Thus productivity trends in the medium run are assumed to be linearly linked to investment and output rates. Moreover, Schumpeterian variables related to technological change are estimated by incorporating them into the constant term a. Similarly, by modifying the traditional accelerator equation, the variation in investment is dependent on consumption and the so-called wage gap, i.e. the difference between productivity and real wages, a very crude proxy for the evolution of the profit share. This change would be detrimental if our aim were to study cycles and stability properties, but seems admissible as far as growth paths are concerned. The last change concerns consumption: profits are assumed to be totally saved, whereas the propensity to consume out of wages is c, not necessarily equal to 1. This is a Kaleckian hypothesis, which does not change too much the global properties of the model.

After making these simplifications, the model is now as follows:

THE BASIC MODEL

(1)	$P\mathring{R}=a+b.\mathring{I}+d.\mathring{Q}$	Productivity equation
(2)	$\mathring{I}=f+v.\mathring{C}+u.(P\mathring{R}-R\mathring{W})$	Investment equation
(3)	$\mathring{C}=c.(N\overset{\circ}{.}RW)+g$	Consumption equation
(4)	$\mathring{R}W=k.P\mathring{R}+l.\mathring{N}+h$	Real wage formation
(5)	$\mathring{Q}=\alpha.\mathring{C}+(1-\alpha).\mathring{I}$	Accounting identities
(6)	$\mathring{N}\approx\mathring{Q}-\mathring{P}R$	Accounting identities

ENDOGENOUS VARIABLES: 6 *PR, I, Q, C, RW, N.*

EXOGENOUS VARIABLES: none, since exogenous factors are reflected by the constant terms a, f and h.

CONDITIONS ON PARAMETERS: $b\geq 0$, $d\geq 0$, $v\geq 0$, $u\geq 0$, $0\leq c\leq 1$, $k\geq 0$, $l\geq 0$, $0\leq\alpha\leq 1$.

However simplistic, this framework is rich enough to deal with most of the issues discussed in the previous section. Let us make a list of the various polar cases:

- Technical progress is not only defined by the exogenous trend a—which might be high or low according to long-run evolution—but by various mechanisms related to productivity formation. These could be due either to capital deepening (high b, possibly $d=0$) or to pure increasing returns to scale, without any link with investment ($b=0$, d important). The last case reflects a learning-by-doing effect 'à la Arrow/Wright'.
- Contemporary macroeconomic theory is highly concerned with the question of investment determination: is investment profit-, demand- or credit-led? The present model deals mainly with the two first alternatives. In the pure Keynesian case, only demand expectations play a role in investment decisions ($u=0$, v related to the capital–output ratio). At the opposite extreme, in the Classical Marxian case, the profit rate is the only factor determining investment ($v=0$, u important).
- Similarly, wage formation is at the core of many discussions about the links between economic policy and the rigidity or flexibility of labour markets. In this model, and contrary to the traditional presentation, the indexation of wages to productivity is seen as the outcome of a Fordist compromise, and not that much as due to pure market mechanisms (in that case, $l=0$, k is positive or even close to one, with $h \approx 0$). In the other extreme case, wage formation is purely competitive, i.e. linked to the evolution of employment, for a given trend in total labour force ($k=0$, l positive and high).

The task then is to play with this small model by first studying in the next section its analytical properties, and then in the following section looking for interpretations of very stylized historical facts.

Very contrasted growth or crisis configurations may exist

For simplicity, the solution of the model can be organized around a very simple idea: where does the productivity growth trend come from? A subpart of the model can be solved in order to associate such a trend with any given level of growth (relation I). Second question: how are productivity gains shared between wages and profits? Another part of the model (of course, with some common equations) allows us to compute the growth in demand associated with each productivity rate (relation II). Let us briefly study these two parts.

The productivity regimes: the joint result of technology, demand and income distribution

Using equations (1), (2), (3) and (4), one gets the following 'sub-reduced form' for productivity:

$$\text{(I)}\quad P\mathring{R}=\frac{b[vc(1+l)-ul]+d}{1-b(vc-u).(k-1-l)}\,.\,\mathring{Q}+\frac{a+bf+vg+b(vc-u).h}{1-b(vc-u)(k-1-l)}$$

i.e.

$$P\mathring{R}=B.\mathring{Q}+A$$

One recognizes the usual form of the so-called Kaldor–Verdoorn relations, but the matter is more complex than initially thought by these two authors.

First, such an equation is not purely a matter of technology, since demand formation and income distribution do play a role. The relation is a function of purely technical aspects if and only if investment has no influence upon productivity ($b=0$), i.e. only in a very special case.

But even if the technical frontier is kept constant, the productivity-growth reduced form can shift or rotate when income distribution changes. For example, when wage earners benefit from more favourable shares (increase in k), the elasticity of productivity with respect to growth increases when investment is more sensitive to demand than to profit ($v/u > 1/c$), and decreases in the opposite case. This last situation might apply to the tendencies observed after 1973, which are characterized by a significant shift in income distribution. This is a possible explanation of the instability stressed by Rowthorn (1975), which seems to be confirmed by the breaking down of the formula since 1973 (Michl, 1984; Boyer and Ralle, 1986a).

Similarly, any change in the investment function shifts the Kaldor–Verdoorn relation in a rather complex manner, since it depends on the whole set of parameters. For low indexation to productivity (i.e. $k<1+l$), a strengthening of the profit motive in investment (rise of u) and milder accelerator effects reduce the apparent size of increasing returns to scale. This is the kind of evolution which seems to have been observed during the last decade.

Combining the factors characterizing technology, investment determination and income distribution gives a large set of configurations. Instead of presenting all of them, it is convenient to discuss only major cases (Figure 27.1). It turns out that four very contrasted cases might appear, according to the intensity of the various mechanisms. Let us stress only two major findings. First, in the pure Fordist case, the usual Kaldor–Verdoorn relation is observed only if wage indexing is not too high (case 3), since up to a certain threshold a perverse configuration may emerge (case 4). Second, the same *ex post* reduced form may result from very different mechanisms: purely competitive wage and exhilarationist effects of profit upon investment (case 2), whereas for more likely hypotheses productivity increases are lower, the higher the growth (case 1).

The previous discussion enables us to present a taxonomy of the various technological systems and/or accumulation regimes. The former can be distinguished according to the configuration of the whole set of parameters

Figure 27.1 The various productivity regimes

I. The pure classical cases

They can be characterized by three principal hypotheses:

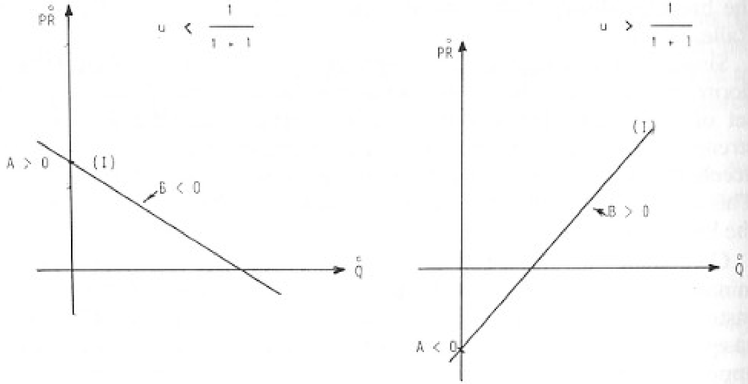
(i) No increasing returns to scale: $d = 0$
(ii) Investment is purely profit driven: $v = 0$ $u \gg 0$
(iii) No *ex ante* productivity-sharing but competitive mechanism for wages: $k = 0$, $l \geq 0$

$$\text{Then by (I): } \dot{PR} = \frac{-b.u.l.}{1-u(1+l)} . \dot{Q} + \frac{a+b(f-uh)}{1-u(1+l)} = B.\dot{Q} + A$$

According to the relative size of the effect of profit upon investment and employment upon real wages, two polar cases can be observed.

Case 1: The profit effect is limited

Case 2: The profit effect is important



The productivity-growth relation is downward sloping

The productivity-growth relation is upward sloping

In both cases we assume that: $u < \dfrac{a+bf}{bh}$

Figure 27.1 The various productivity regimes

II. The pure Fordist cases

By contrast, they are defined by inverting the three hypotheses of the classical case:

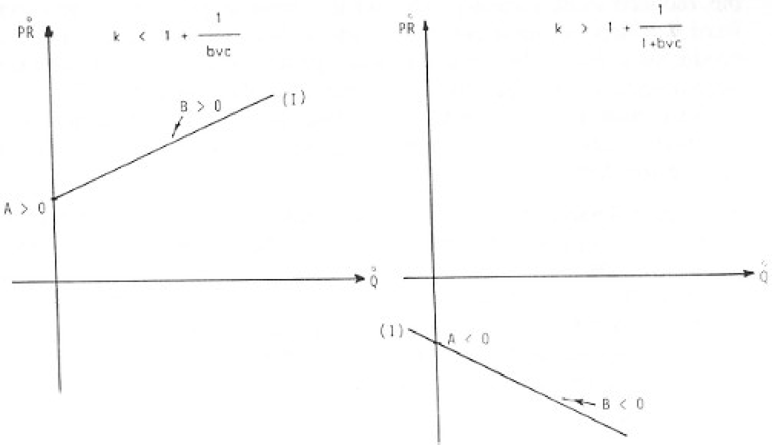
(i) Significant increasing returns to scale: $d \gg 0$
(ii) Investment is totally demand driven: $v \gg 0$ $u = 0$
(iii) Productivity-sharing of wages, absence of competitive mechanism: $k \gg 0$ $l = 0$

$$\text{Then by (I): } \dot{PR} = \frac{bvc+d}{1-bvc(k-1)} . \dot{Q} + \frac{a+bf+vg+bvch}{1-bvc(k-1)}$$

By hypothesis $bvc + d > 0$ and it may be assumed that $(a + bf + vg + bvch) > 0$. According to the relative size of the wage indexing and accelerator effects, two other polar cases can be observed:

Case 3: Wage indexing is limited

Case 4: Wage indexing is high



The usual upward-sloping Kaldor–Verdoorn relation appears as a reduced form

A perverse and negative relation between growth and productivity

$(a, b, d, f, v, \ldots)$. The latter would be defined as mainly extensive or intensive by looking at the reduced form for productivity. Intensive accumulation refers to spillover effects from growth to productivity, and not that much to the size of productivity increases *per se*. Thus extensive accumulation will prevail if B is fairly low, intensive accumulation when B is high.

Demand regimes defined by the wage formation and investment functions

Let us now assess the consequences of a given productivity rate upon the growth of demand, since in this model production is always fixed accordingly. Solving the system without using the productivity equation (1) leads to the following reduced form:

$$\text{(II)} \quad \mathring{Q} = \frac{[\alpha c + (1-\alpha) v c - (1-\alpha) u].(k - l - 1)}{1 - [\alpha + (1-\alpha) v].c(1+l) + l(1-\alpha).u} . \mathring{PR}$$

$$+ \frac{(1-\alpha) f + (ch + g)[\alpha + (1-\alpha).v] - h(1-\alpha) u}{1 - [\alpha + (1-\alpha) v].c(1+l) + l(1-\alpha).u}$$

i.e.

$$\mathring{Q} = D.\mathring{PR} + C$$

We will denote by demand regimes the various configurations taken by this reduced form. Basically, whether the curve slopes upward or downward will depend upon two main factors: first, income distribution, i.e. productivity sharing between wages and profits, and, second, the sensitivity of investment to either profit or demand variations (Figure 27.2).

The complete discussion will not be given here. Instead we shall present four polar cases, combining two extreme hypotheses about investment and income distribution.

- A pure classical demand regime associates profit-led investment with mainly competitive wage formation (configuration 1). In this case, productivity increases promote profits, hence investment and effective demand, which enhance employment, therefore consumption, according to a classical virtuous cumulative growth model. As a consequence, demand increases with productivity. The functioning of this regime can be summarized by a very simple causation mechanism:

$$\text{Productivity} \xrightarrow{+} \text{Profit} \xrightarrow{+} \text{Investment} \xrightarrow{+} \text{Employment} \xrightarrow{+} \text{Consumption}$$

- A hybrid classical demand regime combines demand-led investment with the same competitive wage formation mechanism (configuration 2). The previous mechanisms are then reversed: more productivity induces lower wage increases, hence lower consumption, in such a manner that investment is also reduced via an accelerator based upon

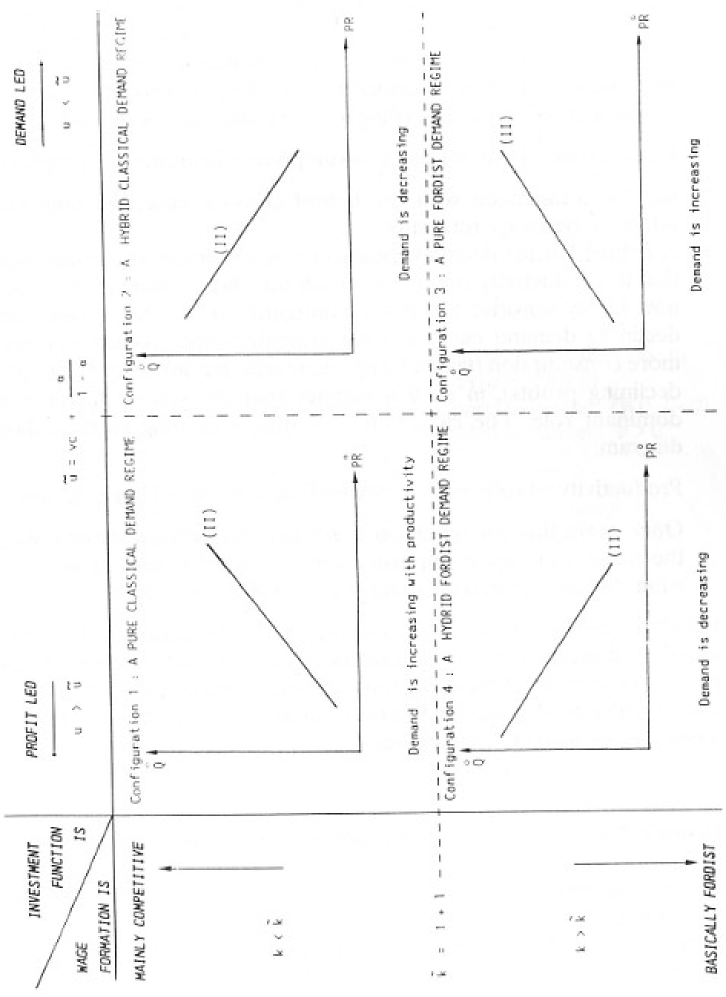


Figure 27.2 The different demand regimes

consumption. Therefore, demand is now declining with productivity, according to the following stylized analysis:

$$\text{Productivity} \xrightarrow{-} \text{Real wages} \xrightarrow{+} \text{Consumption} \xrightarrow{+} \text{Investment} \xrightarrow{+} \text{Employment}$$

- A pure Fordist demand regime associates the same demand-led investment with an explicit sharing of productivity between wages and profits (configuration 3). Now the growth of demand is wage-led: any improvement in productivity raises *ex ante* real wages, hence consumption, investment and effective demand. As in the pure classical case, the law is now increasing but according to a very different mechanism:

$$\text{Productivity} \xrightarrow{+} \text{Real wages} \xrightarrow{+} \text{Consumption} \xrightarrow{+} \text{Investment} \xrightarrow{+} \text{Employment}$$

  But by comparison with the hybrid classical case, the only change relates to real wage formation.

- A hybrid Fordist demand regime can be observed when wage indexation to productivity is up to a certain threshold, whereas investment is now highly sensitive to profits (configuration 4). The rationale of the declining demand curve is clear enough: more productivity induces more consumption (via real wage increases) but less investment (due to declining profits), in such a manner that the second factor plays a dominant role. The causation now runs according to the following diagram:

$$\text{Productivity} \xrightarrow{-} \text{Profit} \xrightarrow{+} \text{Investment} \xrightarrow{+} \text{Employment} \xrightarrow{+} \text{Consumption}$$

  Once again this configuration looks like the hybrid classical one, but the reasons are quite opposite, since the mechanisms of income distribution and demand generation are different indeed.

Finally, a growth (or accumulation) regime can be defined for each combination of productivity-growth relations and demand regimes. Instead of presenting a purely formal analysis, the macro-model will be confronted with the historical trends and periods already examined by the regulation approach, as presented in Chapter 4.

Historical stylized facts over a century: an interpretation

If one considers the evolution of capitalism since the first industrial revolution, the model suggests the possible succession of four stages, combining specific technological systems and forms of socio-economic tuning (Figure 27.3).

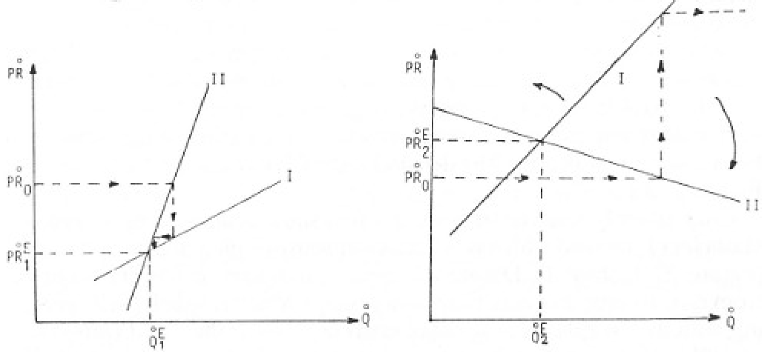
The nineteenth century: moderate increasing returns and investment-led growth

This period is basically characterized by the penetration of new methods of production, via rapid industrial investment promoting this new produc-

Figure 27.3 An interpretation of typical historical periods

Stage 1: The nineteenth century
Moderate but stable growth

Stage 2: The inter-war period
Structural instability and crisis



Hypotheses:

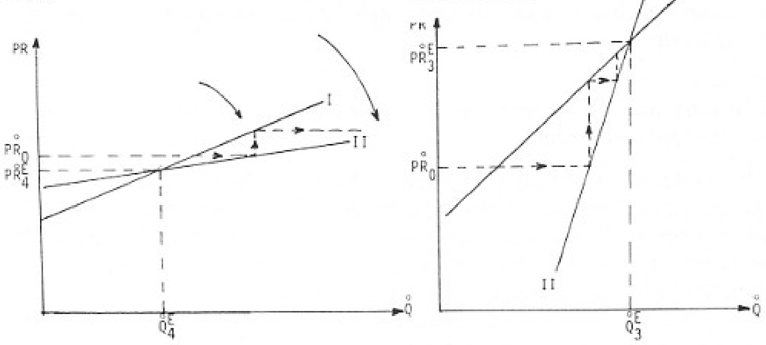
1. Moderately increasing returns of scale ($d \cong 0$, but $b > 0$)
2. Competitive wages (k 0, $l > 0$)
3. Profit-led investment ($v \cong 0$, $u > \bar{u}$)

Hypotheses:

1. Significant returns to scale due to Taylorism ($d > 0$, $b > 0$)
2. Still competitive wages ($k \cong 0$, $l > 0$)
3. Demand-led investment ($v > 0$, $u < \bar{u}$)

Stage 4: The Seventies
Slow and unstable growth: a structural crisis

Stage 3: The 'Roaring Sixties'
Unprecedented high and stable growth



Hypotheses:

1. Exhaustion of increasing returns to scale (d and b declining)
2. An over-indexing of wages to productivity ($k > 1 + l$)
3. The profit motive is back (v is declining, u increasing)

Hypotheses:

1. Fordism brings significant returns to scale ($d > 0$, $b > 0$)
2. Capital–labour compromise over productivity sharing ($k \geqq 0$ $l \geqslant 0$) but $k < 1 + l$)
3. Consumption-led investment ($v \geqslant 0$, $u < \bar{u}$)

tive system. From the technological standpoint, the increasing returns to scale associated with the deepening of the division of labour are in line with the previous trends. Nevertheless, a high ratio of investment allows significant productivity increases. Thus the productivity regime corresponds roughly to a profit-led productivity–growth relationship (as shown by case 2 in Figure 27.5). As far as demand generation is concerned, investment again is the leading factors via an income distribution initially very favourable to entrepreneurs. Wage-earners only benefit from this industrialization process if employment is increasing, wage formation being mainly competitive (the demand regime is close to configuration 1 in Figure 27.2).

Consequently, one can imagine this first stage as associating an average elasticity of demand with a moderately upward-sloping productivity curve (Figure 27.3, stage 1). During this period, industrial crises are observed from time to time, but according to a process which is usually self-regulating, with the exception of the major crisis of 1848 or the Great Depression of 1873–96. It can then be assumed that the growth model is stable, i.e. that any disturbance affecting either productivity or demand levels off. A necessary condition for that property ($|BD| < 1$) is that *ex post* increasing returns and demand elasticity are not too high.

The equilibrium growth path is stable and depends on the accumulation regime. If, for example, innovation raises the exogenous trend of productivity, both productivity and growth rates will be higher. The outcome for employment is related to the elasticity of demand: a reduction if this elasticity is low, an increase if it is high. Similarly, an upward shift usually increases growth and productivity, but the precise results do vary if structural change has effects upon investment, consumption or income distribution.

The inter-war period: a surge of increasing returns and a shift towards demand-led investment

The model allows an interpretation of the contradictions associated with the transition from an extensive to an intensive accumulation regime. In some respects the mode of regulation shows continuities, especially as far as competitive wage formation is concerned. But in other respects two structural changes speed up after the First World War. First, Scientific Management leads to a strengthening of increasing returns, by deepening the division of labour and employing highly specialized equipment. This results in an upward shift of the productivity–growth function (Figure 27.3, stage 2). Second, mass production has to be complemented by mass consumption, so that investment is now linked not only to profits, but also to household consumption. Therefore the demand regime drastically changes, rotating from an upward- to a downward-sloping curve.

The new global regime is then quite new, at odds with the previous one. Of course, medium-term growth rates are logically higher, which explains the unprecedented Roaring Twenties. Potentially, the new technological system accelerates the industrial pace. But the opposite side of the coin has quite dramatic consequences: since the demand regime now declines with productivity, the new technological system ultimately spurs productivity growth but at the expense of employment. Furthermore, the excess of productive capacity over demand is such that the whole system finally becomes unstable. This is a possible theoretical explanation of the vicious spiral of intensive accumulation without explicit mass consumption stressed by previous analyses (Chapter 4 Figure 4.2).

Post-Second World War boom: a capital–labour accord consistent with the technological pattern

After the Second World War the technological paradigm remains more or less the same, but two major structural changes affect income distribution and demand. First, a new capital–labour accord about productivity-sharing induces a brand-new wage mechanism, hence a consumption function increasing with productivity. Second, investment is now more and more linked to demand, and not that much to profit rates, which are already very high. Consequently, aggregate demand becomes an increasing function of productivity, contrary to what was observed during the inter-war period.

The model confirms that there may be a stable and fast growth path within this general accumulation regime with which mass production and consumption are associated (Figure 27.3, Stage 3). The rate of growth is higher, since demand is much more dynamic and spills over to productivity via increasing returns to scale, while capital deepening associated with the accelerator mechanism strengthens even more the productivity–growth relation. The pattern of development is stable if the indexing of wages to productivity is sufficient but not too high, so that it guarantees that any discrepancy between productive capacity and demand is self-correcting (Boyer and Coriat, 1987).

Thus this analysis confirms the previous hypothesis about the shift from an unstable accumulation regime to a stable one during the 1950s. Consequently, the economic system reacts differently to the unfolding of the same technological path. During the inter-war period, more productivity ultimately meant less employment (compare, on Figure 27.3, stage 2 by shifting upwards relation 1). After the Second World War, within the new demand regime, the same movement simultaneously increases productivity, growth and possibly employment (imagine the same shift on Figure 27.3, stage 3).

The present crisis: the exhaustion of the technological path and contradictions over income distribution

The very implementation and diffusion of Fordism set into motion slow adverse trends which finally destroyed the structural stability of the system, and thus made it very vulnerable to external shocks, whether stemming from energy supply or financial markets. Among the three factors to be taken into account, the struggle for external competitiveness cannot be treated within a closed macroeconomic model, but the other two have consequences which are easier to analyse.

The fact that most economies operated at quasi-full-employment level largely benefited wage earners, who at the end of the 1960s won significant increases in real wages and a rise in the degree of indexing, explicitly in terms of consumer prices, implicitly in terms of medium-term productivity gains. Therefore demand becomes more sensitive to productivity, if investment is still buoyant, and even if profit trends might be deteriorating. Above a certain threshold the growth path becomes unstable (Figure 27.3, stage 4), which seems likely given the developments in OECD countries since the early 1970s.

The erratic character of demand aggravates the productivity problem, since markets are more and more difficult to forecast and do not allow the increasing returns associated with stable and growing markets to be realized. But the underlying difficulties of Fordism are much more severe: more capital is needed to get the same labour-productivity growth, and the maturing of the technological system makes it less efficient in improving industrial organization. This second change shifts downward the productivity-growth function (for example, in the United States) or even reduces drastically the significance of the Kaldor–Verdoorn hypothesis (Mitchl, 1984; Boyer and Ralle, 1986a). Consequently, the rate of growth is itself reduced, a second feature of the present crisis.

Of course, this sketch is more suggestive than really demonstrative. Many detailed statistical and econometric studies will be needed in order to support these hypotheses. Preliminary estimates for the United States (Caussat, 1981) or EEC countries (Boyer and Petit, 1986) seem rather promising as regards the general structure of the model, if not the precise timing of the stages (Boyer, 1986a). It is now time to derive some prospective views from this historical perspective.

The present economic and technological transition: is a new growth regime emerging?

It would take us too far afield to present the various and very contradictory transformations now affecting different institutional structures (nature of competition, wage–labour nexus, state interventions, international and monetary regimes). One can find such analyses in recent publications by the authors of the 'régulation school' (Aglietta and Orlean, 1982; Aglietta and Brender, 1984; Boyer, 1986c; Boyer *et al.*, 1987). Here the emphasis will be upon both the technological system and the wage–labour nexus. Among many combinations of different evolutions, two typical scenarios will now be presented which serve as illustrations of the use of the model.

Wage austerity and traditional technical flexibility: towards stagnation?

It may not be an exaggeration to talk about a complete breakdown of the whole pattern of industrial relations typical of Fordism: decentralization of collective bargaining, general de-indexing, guidelines by governments implying stability of real wages, and significant changes in wage differentials. Thus the shift in income distribution observed after 1973 has been reversed during the 1980s, with productivity gains now accruing mainly to profits and far less to wage-earners.

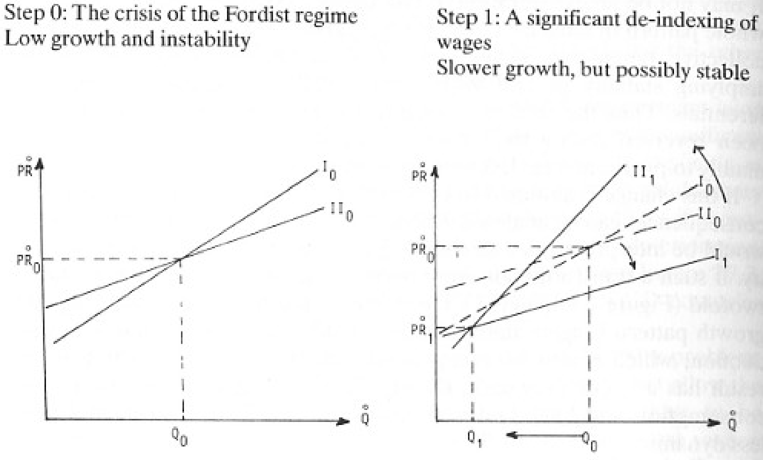
If this change is assumed to be a structural and lasting phenomenon, its consequences can be analysed within the previous Fordist growth model. It would be interpreted as a de-indexing of wages with respect to productivity. If such a transformation were permanent, the consequences would be twofold (Figure 27.4, step 1). Up to some threshold for de-indexing, the growth pattern is again stable since demand is then kept in line with production, which is new by comparison with the 1970s (step 0). But this result has a major drawback: the equilibrium growth rate is lower since consumption and hence investment (via an accelerator mechanism) are less dynamic.

A second change concerns investment determinants. It seems that the econometric equations estimated before the 1980s usually overestimate the recovery of investment for a given increase in aggregate demand (for example, Artus and Muet, 1984). Two different factors might explain this new pattern. First, the atypical configuration of most key macroeconomic variables (level of demand, rate of return, real interest rate) makes firms more cautious before deciding upon investment. Second, the specialists in industrial organization are suggesting that the introduction of electronic devices into industrial processes and services has reduced the bottlenecks associated with earlier Fordist equipment, which was highly specialized. Since the same equipment can be shifted from one product to another of the same variety, the investment level will react slower and to a lesser extent to the same increase in demand (Bultel, 1983; Kundig, 1984).

Whatever the reasons (the macroeconomic ones might be dominant for the 1980s, the flexibility argument possibly more significant in the long run), this change weakens the accelerator mechanisms in the investment function (decrease in parameter v without any offsetting increase in u or f). It is easy to check that demand is now less sensitive to productivity, i.e. that its slope is steeper in the usual (growth–productivity) diagram (Figure 27.4, step 2). But simultaneously, lesser accelerator effects induce lower productivity increases via the traditional capital–labour substitution mechanism. As a consequence, these shifts reinforce the impact of wage de-indexing: the economic system is stabilized but at the cost of a reduction in medium-term growth.

One last hypothesis about the technological system has to be added. Among the very contradictory trends observed in the last one or two decades, some observers stress that the flexibility–productivity trade-off

Figure 27.4 A first scenario: wage austerity and technical flexibility



Step 0: The crisis of the Fordist regime
Low growth and instability

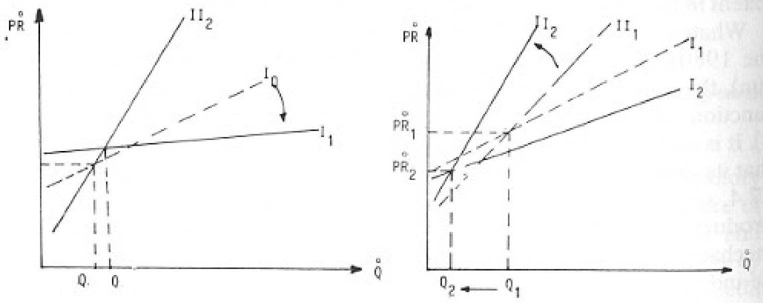
Hypotheses:

1. Low increasing returns and investment efficiency (d and b small)
2. Overindexing of wages to productivity ($k > 1 + l$)
3. A mix of profit and demand in investment (v and u average)

Step 1: A significant de-indexing of wages
Slower growth, but possibly stable

Changes:

2′. Subindexing of wages ($k < 1 + l$ and stability condition)



Step 3: Flexible specialization
Lower and stable growth

Changes:

2′, 3′ and 1′: almost complete flattening of the law of return

Step 2: More investment inertia
Still slower growth, but stable

Changes:

2′ and 3′: lesser accelerator effect (v is lower)

has been shifting towards methods to cope with variability, even at the cost of lower productivity increases. Hence, even if from a static point of view flexible equipment is superior to specialized machines, the cumulative improvement in technical efficiency is inferior. Impressed by Italian industrial organization, some authors have prognosticated a new industrial divide, far from Fordist mass production, towards a modernized and computerized variant of the Proudhonian logic (Piore and Sabel, 1985).

Taking into account that at an aggregate level and for mature industries the Kaldor–Verdoorn relation has broken down, one could expect the complete disappearance of increasing returns and a moderate exogenous increase in productivity near the trend observed in the 1980s. Consequently, the productivity relations would rotate clockwise and become horizontal. Here again, this shift has two opposite effects: lower, but stable growth (Figure 27.4, step 3).

This scenario roughly extrapolates some of the spontaneous developments of the present decade. The puzzling conclusion would be that a series of strategies of adaptation to the crisis might induce, at the macroeconomic level, stability within stagnation, but without eliminating mass unemployment. This hint is linked indeed to a very specific model, and a crucial question remains open: what other mechanisms could counteract these disappointing tendencies?

A compromise about flexible automation and income distribution: difficult to reach, promising for employment and stability

At the core of this second scenario is a general hypothesis about the novelty of socio-technical trends observed during the last decade. The Fordist conception assumed a clear-cut distinction between productivity gains generation (originating in scientific management and specialized equipment) and their sharing via collective bargaining. Nowadays, in many instances, such a split is detrimental both to economic efficiency and workers' expectations, so that key bargaining—explicitly or more often implicitly—now revolves around know-how, motivations concerning productivity, and quality. Therefore a possible new compromise compatible with the present process of industrial restructuring would be more readily accepted if firms concentrated on defining new wage formation guidelines leading to a sharing of the benefits.

From the technological standpoint, the key issue is about the possible productivity regime associated with such a new New Deal. According to a rather widely accepted view, economies of scope would replace economies of scale and therefore increasing returns to scale would no longer be at the core of the competitive mechanism. Productivity trends would then be quite independent of growth, even if potentially large. Actually, a lot of evidence suggests a more balanced view. First, it can be shown that economies of scope might be complementary to economies of scale (Bailey and Friedlander, 1982), since the same inputs, equipment and know-how can be shared by various products. Second, a significant increase in final product variety can be obtained by combining different, highly standardized subparts; therefore differentiation and economies of scale can be jointly reaped. Third, detailed studies of experience curves show that the more recent goods (disc memory drives, digital watches, integrated circuits, MOS dynamic ram, etc.) are even more sensitive to such effects than typical Fordist goods (model-T Ford, steel production, etc.) (Ayres, 1985). Finally, such a diagnosis seems to be confirmed by a purely macroeconometric study of productivity regimes since 1973 (Boyer and Coriat, 1987). The industries in which demand is buoyant are experiencing very significant returns to scale (between 0.7 and 0.8).

Therefore a future upward shift of the productivity regime is likely to the extent that the new technological system is implemented in new industries and possibly some modern services. This would exert a positive influence upon medium-term rates of growth and productivity increase. Given an adequate elasticity of demand, this would promote a recovery of employment (Figure 27.5, step 1 compared with the initial equilibrium in step 0). It has to be noted that this tentative new virtuous circle presupposes that demand increases with productivity.

This feature is indeed crucial. Since demand characteristics matter even in the medium or long run, it is possible to design a capital–labour accords in order to satisfy two different objectives: speeding up growth, without generating structural instability. Given all other parameters of the economy and technology, productivity-sharing has to be bounded by two limits (Boyer and Coriat, 1987). In such a case, the configuration of the system is very favourable: industrial modernization and job creation might again be coherent (Figure 27.5, step 2). From a purely economic standpoint, the economic system is very close to the typical Fordist one (Figure 27.3, stage 3). Nevertheless, from a social and technological viewpoint, the outlook is quite different as regards work organization, the nature of products, and the structure of industrial relations and collective bargaining.

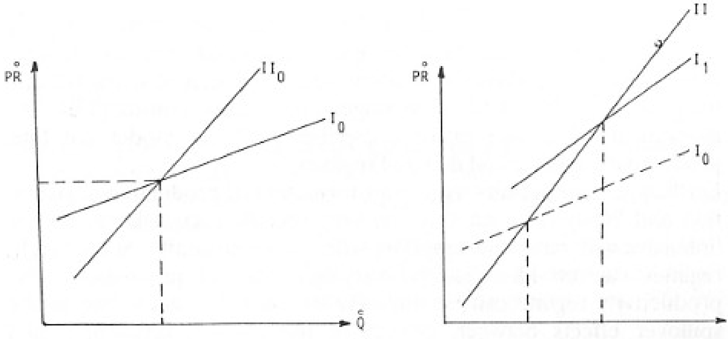
Concluding remarks

The present chapter has tentatively combined two lines of analysis: on the one hand, the 'régulation approach' which stresses the succession of various accumulation regimes, and, on the other, a renewal of a post-Keynesian theory of growth. In comparison to previous work, some steps have been made towards a better integration of macro modelling and historical analysis. First, a whole family of macro models has been proposed in order to substantiate the basic hints of Chapter 4. Second, the hypothesis of full employment, often made in the 1960s by post-Keynesians, is removed in order to deal with cases in which the labour force and employment are diverging. Third, the cumulative causation model proposed by

Figure 27.5 A second scenario: a cooperative approach to automation and income distribution

Step 0: The initial situation
Stability within stagnation

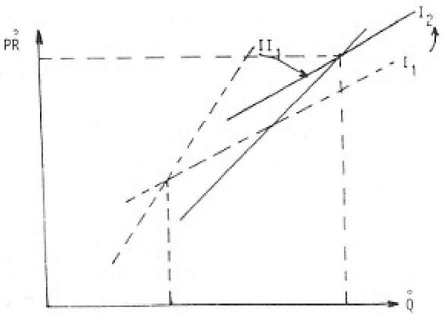
Step 1: A surge in increasing returns
More growth, and possibly employment



Hypotheses: same as Step 2, Scenario 1

Changes: 1′. Faster technical change (*d* higher, or *a*)

Step 2: An adequate productivity-sharing
A possible way out of the crisis



Changes: 2′. Higher share of productivity to wage earners (*k* high but such that $BD < 1$)

Kaldor (and especially the so-called Kaldor–Verdoorn equation) is analysed within a more detailed and precise framework, in order to explain possible structural shifts in the model. Fourth, special attention has been devoted to the consequences of the matching or mismatching between the technological regime and institutional conditions (Chapter 3).

Four major results emerge from the analysis:

1. One of the main features of the theory is its roots in the following question: how are productivity gains generated (via innovations, capital deepening, division of labour and extension of markets), and how are they shared between wages and profits, consumption and investment? The two major components of the model are thus productivity regimes and demand regimes.
2. Fordism, i.e. the simultaneous transformation of productive organization and lifestyle, appears as one very specific accumulation regime (intensive with mass consumption) within a whole family. At least eight regimes can be identified by varying structural parameters. The productivity regime can be intensive or extensive, according to the spillover effects between growth of the market, investment and productivity. Similarly, four demand regimes were found: pure classical, hybrid classical, Fordist, and hybrid Fordist, according to the relative importance of demand and profits in investment decisions and the nature of the wage formation process.
3. Further analysis shows that only some of these regimes are viable, i.e. they induce a stable growth path. The technological and industrial characteristics at the origin of the productivity regime have to be compatible with the income distribution mechanism (either competitive, 'monopolist' or 'Fordist') and demand generation (relative size of consumption and investment, profit or demand-led investment). Some subperiods of the nineteenth century and still more the post-Second World War long boom are examples of such a virtuous circle.
4. Conversely, structurally unstable systems might correspond to structural crises (for example, the 1929 crash) or depressions, sometimes called phase B in the Kondratiev waves literature. Therefore this is a possible reinterpretation of these findings. During the first period, the very success and deepening of a given accumulation regime induces a slow shift in structural parameters (upswing of phase A). But beyond some threshold (given by the condition of stability), the system becomes unstable, which leads to the crisis of the whole technological system and the institutional forms. According to this view, the way out of structural crisis would be neither automatic nor deterministic, but would depend upon innovations, social and political struggles, trials and errors, as well as chance.

Of course, all these findings are not definitive at all, since they rely upon a very specific macroeconomic model. Therefore further analysis is definitely needed. From an empirical standpoint, the stylized historical facts should be confronted with the results of very detailed studies. Similarly, the origins of the present crisis must be investigated again, in order to check whether this description fits the simulations of a full fledged macroeconomic model. From a theoretical point of view, the model of the economy should be opened to external trade and capital flows. Furthermore, a careful study of dynamic patterns (especially of long- and medium-term cycles) should be integrated into the analysis of growth trends.

References

Aglietta, M. (1974), 'Accumulation et régulation du capitalisme en longue période: exemple des États-Unis (1870–1970)', Thèse Paris I, Octobre.

Aglietta, M. and Brender, A. (1984), *Les Métamorphoses de la société salariale*, Paris, Calmann-Levy.

Aglietta, M. and Orléan, A. (1982), *La Violence de la monnaie*, Paris, P.U.F.

Artus, P. and Muet, P. A. (1984), 'Un panorama des développements récents de l'économétrie de l'investissement', *Revue Économique*, vol. 35, No. 5, Septembre.

Ayres, R. U. (1985), 'A Schumpeterian model of technological substitution', in *Technological Forecasting and Social Change*, vol. 27, Elsevier Science Publishing Company.

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

Bertrand, H. (1978), 'Une nouvelle approche de la croissance française de l'après-guerre: l'analyse en section productives', Statistiques et Études Financières, Série Orange, No. 35.

—— (1983), 'Accumulation, régulation, crise: un modèle sectionnel théorique et appliqué', *Revue Économique*, vol. 34, No. 6, Mars.

Billaudot, B. (1976), 'L'accumulation intensive du capital', Thèse Paris I.

Boyer R. (1975), 'Modalités de la régulation d'économies capitalistes dans la longue période: quelques formalisations simples', CEPREMAP, Juin, ronéotypé.

—— (ed.) (1986a), *La Flexibilité du travail en Europe*, Paris, La Découverte.

—— (ed.) (1986b), *Capitalismes fin de siècle*, Paris, Presses Universitaires de France.

—— (1986c), *La Théorie de la régulation: une analyse critique*, Paris; La Découverte (American edition, Columbia University Press, 1988).

—— (1986d), 'Informatisation de la production et polyvalence…ou comment une flexibilité peut en cacher une autre', *Formation, Emploi*, No. 14, *La Documentation Française*, Avril–Juin.

—— (1987), 'Réflexions sur la crise actuelle', *Revue Française d'Économie*, No. 5, Mai.

Boyer, R. *et al.* (1987), 'Aspects de la crise', Rapport de Recherche CGP/CEPREMAP No. 23–84, Réf, 4584/60-01, 3 Tomes, Février.

Boyer, R. and Coriat, B. (1987), 'Technical flexibility and macro stabilisation: some preliminary steps', *Ricerche Economiche*, No. 4, 1986.

Boyer, R. and Mistral, J. (1978), *Accumulation, inflation, crises*, Paris, P.U.F. (2ème edition, 1983).

Boyer, R. and Petit, P. (1981a), 'Employment and productivity in the EEC', *Cambridge Journal of Economics*, vol. 5.

— (1981b), 'Forecasting the impact of technical change on employment: methodological reflections and proposals for research', in Diettrich and Morley (eds), *Relations Between Technology Capital and Labour*, Commission of European Communities EUR 8181 EN, pp. 49–69.

— (1981c), 'Progrès technique, croissance et emploi: un modèle d'inspiration Kaldorienne pour six industries européenes', *Revue Économique*, vol. 32, pp. 1113–53.

— (1984), 'Politiques industrielles et impact sur l'emploi: les pays européens face à la contrainte extériure', *Revue d'Économie Industrielle*, No. 27, Janvier.

Boyer, R. and Ralle, P. (1986a), 'Croissances nationales et contrainte extérieure avant et après 1973', *Economie et Société*, Cahiers de l'ISMEA, No. P29.

— (1986b), 'L'insertion internationale conditionne-t-elle les formes nationales d'emploi? Convergences ou différenciations des pays européens', *Economie et Société*, Cahiers de l'ISMEA, No. P29, Janvier.

Bultel, J. (1983), 'Flexibilité de production et rentabilité des investissements: l'exemple de la robotisation de l'assemblage tôlerie en soudage par points', *Revue d'Économie Industrielle*, No. 26, 4ème trimestre.

Caussat, L. (1981), 'Croissance, emploi, productivité dans l'industrie américaine (1899–1976)', CEPREMAP, Septembre, mimeograph.

Fagerberg, J. (1984), 'The "régulation school" and the classics: modes of accumulation and modes of regulation in a classical model of economic growth', Couverture Orange CEPREMAP, No. 8426.

Kaldor, N. (1966), *Causes of the Slow Rate of Growth of the United Kingdom*, Cambridge, Cambridge University Press.

— (1975), 'Economic growth and the Verdoorn law: a comment on R. Rowthorn's article, *Economic Journal*, vol. 85, pp. 891–6.

Kundig, B. (1984), 'Du taylorisme classique à la flexibilisation du système productif: l'impact macroéconomique des différents types d'organisation du travail', *Critiques de l'Économie Politique*, No. 26–27, Janvier–Juin.

Mitchl, Th. (1984), 'International comparison of productivity growth: Verdoorn's law revisited', Colgate University, Hamilton, NY, July, mimeograph.

Piore, M. and Sabel, Ch. (1985), *The Second Industrial Divide: Possibilities of Prosperity*, New York, Basic Books.

Rowthorn, R. (1975), 'What remains of Kaldor's law?', *Economic Journal*, vol. 85, pp. 10–15.