EE

APPLIED EVOLUTIONARY ECONOMICS AND COMPLEX SYSTEMS

Edited by John Foster and Werner Hölzl

Applied Evolutionary Economics and Complex Systems

Applied Evolutionary Economics and Complex Systems

Edited by

John Foster

Professor of Economics and Head of the School of Economics, The University of Queensland, Brisbane, Australia

and

Werner Hölzl

Research Group Growth and Employment in Europe and Department of Economics, Vienna University of Economics and Business Administration, Vienna, Austria

Edward Elgar

Cheltenham, UK • Northampton, MA, USA

Published by
Edward Elgar Publishing Limited
Glensanda House
Montpellier Parade
Cheltenham
Glos GL50 1UA
UK

Edward Elgar Publishing, Inc.
136 West Street
Suite 202
Northampton
Massachusetts 01060
USA

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication Data

European Meeting on Applied Evolutionary Economics (2nd: 2001: Vienna University of Economics and Business Administration)
Applied evolutionary economics and complex systems / edited by John Foster and Werner Hölzl
p. cm.
'Selected papers presented at the second European Meeting on Applied Evolutionary Economics (EMAEE), held at the Vienna University of Economics and Business Administration in September 2001'—Pref.
Includes bibliographical references and index.
1. Evolutionary economics—Congresses. 2. Complexity (Philosophy)—Congresses. 3. Economics—Congresses. I. Foster, John, 1947– II. Hölzl, Werner, 1969– III. Title.

HB97.3.E84 2001
330.1—dc22 2004040594

ISBN 1 84376 369 9

Printed and bound in Great Britain by MPG Books Ltd, Bodmin, Cornwall

Contents

Contributors

Jeroen C.J.M. van den Bergh, Department of Spatial Economics, Faculty of Economics and Business Administration and Institute for Environmental Studies, Free University of Amsterdam, The Netherlands

Franco Bevilacqua, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy and MERIT, Maastricht University, Maastricht, The Netherlands

Bernd Ebersberger, VTT, Technical Research Center of Finland, Group for Technology Studies, Espoo, Finland

John Foster, The School of Economics, The University of Queensland, Brisbane, Australia

Koen Frenken, Department of Economic Geography and Urban Research Centre Utrecht (URU), Utrecht University, The Netherlands

Thomas Grebel, Department of Economics, University of Augsburg, Germany

Horst Hanusch, Department of Economics, University of Augsburg, Germany

Werner Hölzl, Research Group Growth and Employment in Europe and Department of Economics, Vienna University of Economics and Business Administration, Vienna, Austria

Jackie Krafft, CNRS Research Fellow at IDEFI-LATAPSES, University of Nice–Sophia Antipolis, France

Alessandro Nuvolari, Eindhoven Centre for Innovation Studies (ECIS), Eindhoven University of Technology, Eindhoven, The Netherlands

Michael Peneder, Austrian Institute of Economic Research (WIFO), Vienna, Austria

Andreas Pyka, Department of Economics, University of Augsburg, Germany

Andreas Reinstaller, MERIT, Maastricht University, Maastricht, The Netherlands and Department of Economics, Vienna University of Economics and Business Administration, Vienna, Austria

Deborah Tappi, Max Planck Institute for Research into Economic Systems Evolutionary Economics Group, Kiel, Germany

Adriaan van Zon, MERIT, Maastricht University, Maastricht, The Netherlands

Preface

This book has evolved from selected papers presented at the second European Meeting on Applied Evolutionary Economics (EMAEE), held at the Vienna University of Economics and Business Administration in September 2001 around the theme 'Advancing Empirical Research Methodologies in Evolutionary Economics'. The chapters deal with various aspects of applied evolutionary economics and related policy issues. These contributions are at the frontier of the field, and selected to focus upon key areas currently under investigation. There is no attempt to cover all aspects of current research. The book reflects the fact that evolutionary economics has reached a new level of maturity as it develops coherent empirical strategies and moves beyond a narrowly construed reliance on models of competitive selection towards a more general analytical framework built upon complexity theory. The book will be useful to researchers interested in economic evolution, economic growth, economic development, innovation, industrial organization and modelling economic evolution. It will also be useful to advanced students in all of these fields.

We would like to thank all those who made the conference, and therefore also this book, possible. First, we would like to thank Koen Frenken and Andreas Pyka for proposing the idea of an EMAEE conference in Vienna. Thanks are due to the Department of Economics and the interdisciplinary Research Group Growth and Employment in Europe, both at the Vienna University of Economics and Business Administration, for organizing the conference. Finally, we would like to thank the following organizations for their material support, without which this book would not have come into existence: the Austrian Ministry for Transport, Innovation and Technology (BmVIT), the Austrian National Bank (ÖNB) and the Department of Cultural Affairs, City of Vienna.

1. Introduction and overview

John Foster and Werner Hölzl

Modern evolutionary economics is just over two decades old and its research programme continues to expand in innovative directions. A key development in recent years is the move away from the traditional focus on processes of selection towards the dynamics of complex systems (Foster and Metcalfe 2001). Increasingly, the economy and its components are seen as complex adaptive systems that change qualitatively in historical time. This shift in perspective has led to suggestions that new analytical tools and different empirical research methodologies are necessary. In the natural sciences, approaches to the development and functioning of systems, which apply both evolutionary and self-organization theory in a consistent and compatible way, have proven to be fruitful in recent years (see Kauffman 1993, Depew and Weber 1995). However, the economy, as a system with component sub-systems, is different in important ways to those found in physio-chemical and biological contexts. The challenge is not only to discover analytical representations of economic evolution but also to be able to connect them with empirical research.

New and innovative methodologies and methods of empirical research are being developed by applied evolutionary economists. It is these that are of primary interest in this volume. Consequently, it is not the purpose here to review the literature on modern evolutionary economics, nor to provide a systematic account of empirical research methodologies that are still in their infancy. What is offered in the chapters that follow is, necessarily, selective, exploratory and illustrative. However, before providing an overview of these contributions, it is worth giving a short summary assessment of the connections between evolutionary economics and complex systems theory.

1. EVOLUTIONARY ECONOMICS

Evolutionary economists see the economy as a scientific domain characterized by non-equilibrium processes in which economic agents create and adapt to novelty through learning rather than as a system in which there are disequilibrating shocks to stable equilibrium states (see Witt 1993, Nelson 1995, Saviotti 1996, Foster and Metcalfe 2001, Fagerberg 2003). In the natural sciences, equilibrium is either a state that exists in a completely connected force field (Mirowski 1989) or it is a state of maximal disconnection, for example thermodynamic equilibrium. The former concept of equilibrium is used to address the real world through the application of experimental controls which localize such a force-field equilibrium in space and time. Neoclassical economists have adopted this kind of conception of equilibrium but reversed its role, whereby a set of controls (assumptions) is proposed in a logical setting and analytical experiments are conducted that involve the successive relaxation of these controls. This, of course, does not necessarily get us closer to reality.

The latter case of equilibrium is quite realistic but empty as an analytical tool – it is something to which systems tend in the presence of processes such as entropy. This is a form of equilibrium towards which systems try to resist moving, and it is this resistance and the associated flows of energy and information that self-organization theory seeks to address. Such a vision of the dynamics of systems of the economic kind was grasped in an intuitive way by Joseph Schumpeter (Foster 2000). Correspondingly, the contributions in this volume are strongly influenced by his work.

Schumpeter considered economic evolution as an open-ended process of qualitative change. Evolutionary economics, as it is conceived of in this volume, takes his line of reasoning. In this sense it is different from other strands of economics that use the term 'evolutionary', as, for example, in evolutionary game theory. Evolution, in the context of this volume, is economic development over time and is an open-ended dynamic process over an open state space. Although there are overlaps with evolutionary biology, the Schumpeterian perspective is quite distinct. However, some general principles are shared, for example, replicator dynamic processes which have provided a core for modern evolutionary economics following on from the contribution of Nelson and Winter (1982).

Building on this seminal contribution, modern evolutionary economics came to be based upon the interaction of processes of behavioural variation in a population of heterogeneous economic agents characterized by a certain degree of inertia (heredity), selection and replication. Accordingly, for most evolutionary economists competition is a process of successive selection, where the least efficient products and firms are driven out of the market. What makes a firm competitive are its technological and organizational attributes, which make it worse or better adapted to a specific economic environment, and its capability to learn and innovate, which allows it to improve its position in relation to other firms in the market.

The adoption of replicator dynamics has been crucially important in establishing modern evolutionary economics as a credible and coherent force in economics. However, the strength of evolutionary economics does not rest exclusively on the application of these principles to the economic process. It also lies in the radical change in theoretical perspective that it has introduced into economics, away from concepts such as representative agents and individual optimization and an associated focus on understanding the properties of a possible equilibrium states towards a search for explanations of structural change in the economy. There is no one-to-one transfer of evolutionary analogies and metaphors from biology to economics.

Evolutionary economists take into account specific aspects of the economic domain that are not relevant in the biological domain. In our view the three distinguishing and interrelated traits of evolutionary economics are:

1. While there is disagreement on the specific definitions, there is agreement that *knowledge and information* are central ingredients of the approach of evolutionary economics. Economic systems are knowledge-based. The primary interactions are exchanges between knowledge as a structure and information as a flow. Economic knowledge resides in the mind, both individually and collectively, and leads to the formation of rules, routines and other institutions that facilitate economic coordination. These are reinforced and reproduced through practice. If rules are unused they cease to be relevant and the associated knowledge disappears. It is the processes of knowledge creation and destruction that underpins and drives economic growth and qualitative change. The growth of knowledge cannot be meaningfully captured as a constellation of equilibrating forces (Metcalfe 1998, Foster and Metcalfe 2001, Fagerberg 2003).
2. Evolutionary economics takes a *population approach* instead of a typological approach based on representative agents. The heterogeneity of economic behaviour is based on the distribution of knowledge within the economy. The division of labour is deeply related to the division of knowledge (Hayek 1945). Heterogeneity drives economic change, which can be cast in terms of observable changes in the composition of populations of firms, technologies and industries. The economic system contains a large number of heterogeneous agents that act simultaneously. The interaction between economic agents takes the form of competition and cooperation. Together with spillovers, the decentralized organization creates not only the problem-solving capability of the economic system but also the capability to formulate new problems and new behaviour. Thus the link between aggregate growth and the

transformation of economic structures cannot be captured by employing a typological approach based on representative agents (Metcalfe et al. 2003).

3. The *interdependence between selection and development* is a primary characteristic of evolutionary economics. Competition as a replicator dynamic process structures economic activity (Metcalfe 1998) and selects the most productive techniques, the best organizational arrangements, the most capable human capital and the most attractive products. From the perspective of variety generation, markets are institutions that not only coordinate economic behaviour but facilitate change, entrepreneurship and challenges to established behaviours. Selection processes destroy variety. The generation of variety and the selection of variety interact in the process of development. In order to have economic development, variety needs to be constantly re-created. It is variety in the knowledge structure that permits the novel ideas that result in organizational and technological innovations to emerge.

Economic systems are subject to developments characterized by qualitative, structural and irreversible change. Although such systems can approach steady or even stationary states, these are not equilibrium states of the conventional type. Evolutionary economists do not see economics as a science that is primarily concerned with the analysis of the best use of given resources, that is, optimal outcomes, but, instead, seek analytical representations of the processes of consolidation and change in economic systems. In this respect, evolutionary economics revisits many of the questions asked by the classics (especially Adam Smith and Karl Marx), while the theme of qualitative change in the economic system raises issues considered in a largely intuitive manner by, for example, Alfred Marshall and Joseph Schumpeter, using new theoretical and empirical tools. The strong focus on technological and organizational change within evolutionary economics reflects these historical connections.

However, it has become apparent to many evolutionary economists that perceiving the economy as a set of interacting non-equilibrium processes that exhibit homeostatic tendencies requires a fundamental rethink as to how the complicatedness that we observe in the data can be simplified in a way that permits analysis. The mainstream favours simplistic, rather than simple, representations of the complicated real world and, indeed, such an approach is also favoured in some contributions in heterodox economics. There is a growing realization that, if the complicatedness that we observe is, in fact, due to the behaviour of interacting complex systems, then simple representations of the economy can be derived for the purposes of analysis (Foster 2003). So let us now examine what is meant by a 'complex system' and why this is fundamentally important in evolutionary economics.

2. COMPLEX SYSTEMS

The complicatedness of reality is a daunting prospect for anyone in search of tractable analytical principles. The great task is to provide a universally valid set of principles that can offer simple representations of the structures and processes in the real economy – not principles that avoid confronting this reality but those that acknowledge the fact that we are dealing with complex systems and related complex processes. By definition, such as set of principles is orthogonal to conventional ones as embodied, for example, in Walrasian general equilibrium theory. Such a theory is a simplistic, that is, not an analytical, representation of real processes since it deals with outcomes, not historical processes (Dopfer 2001).

Complex systems in the economy are, at the same time, complicated and organized in a way that permits them to absorb energy and information to create both physical and knowledge structures that allow them to maintain their structural integrity and to develop. Such systems can only be understood in a historical continuum and, as such, can be analysed in simple (not simplistic) ways that is not just historiography. The reduction of complexity to simple analytical representations is not an exercise in mathematical deduction but, rather, a question of understanding the endogenous tendencies that complex systems display and how such tendencies interconnect with those of other systems and the rules that arise from social, political and cultural origins. Applied evolutionary economics is concerned with the manner in which the unique dimensions of these interconnections in particular cases relate to general processes that complex systems all exhibit.

Thus the dynamics of complex systems are, in part, explainable and predictable in so far as they behave on average in a representative (average) manner, and, in part, inexplicable and unpredictable, as they act also in non-representative and non-ergodic ways. A defining characteristic of complex systems is the interdependence of the elements that make up any system. A system containing a number of elements cannot be reduced to systems with a lower number of elements without changing the defining character of the system since the connective structure will be altered (Potts 2000). Complex systems are only partially decomposable. Thus the quest for simple analytical principles is quite different from the conventional reductionist approach. What is sought is an understanding of the temporal and spatial patterns exhibited by complex systems that can be represented in a simple and tractable manner – this lies at the heart of all evolutionary economic analysis.

All this is in stark contrast to mainstream approaches that place the optimizing behaviour of economic agents at the core of economic analysis. Of course, this does not deny that economic agents try to optimize; the crucial issue is the context in which such behaviour occurs. If the economic system is a complex and open system, this has fundamental implications for the behavioural characteristics of economic agents. In the face of the complicated geometry of economic interactions (Potts 2000), economic problems are computationally demanding. So economic agents must be viewed as having limited computational power. In trying to explain phenomena in historical reality, they cannot be just viewed generally as optimizing agents with perfect information or rational expectations. In a complicated reality inhabited by complex systems, rationality becomes a procedural and limited notion, or as Herbert Simon called it, bounded rationality. Such rationality manifests itself when economic agents attempt to solve problems in economic processes directed towards goals or aspirations that need not be rationally constructed because they are formed in uncertain circumstances.

The implications of computational difficulty are not confined to the behavioural attributes of economic agents. Such difficulty also places inherent limits on the ability of economists to model the behaviour of complex systems. It is striking just how quickly conventional 'simplistic' linear models of economic behaviour become mathematically complicated and intractable simply by introducing an element of non-linearity. Complicated mathematics is generally unhelpful in understanding the behaviour of economic systems. The complex systems perspective involves recognition that there is a large self-organizational element in the behaviour of economic systems over time that can be summarized in mathematical terms, that is, in terms of growth rates. Thus the behaviour of complex systems can be simply represented in terms of tendencies that are punctuated by exogenous impacts. The theory that lies behind such tendencies is concerned with the emergence of rules, routines and other connections between individuals and groups that result in economic value (Dopfer et al. 2004).

The rule-based interactions of agents result in emergent features at the macroeconomic level of the economy. The challenge for evolutionary economists is not to prove the stability or existence of equilibrium in the economic system, but to explain and understand which rules govern the interaction between the agents and which processes and interactions change the rules of the economic game over time, enabling the system to evolve. Even if a complex economic system cannot be exactly decomposed into modular independent subsystems, the notion of nearly decomposable systems allows complexity to be analysed in systematic ways. Thus components of the economic system can be dealt with as self-organizing entities where replicator dynamic representations can be used to capture evolutionary tendencies both within such components and between them (Foster and Metcalfe 2001). However, much has to be done before a unified and simple analytical framework, built from complexity theory, can emerge in evolutionary economics. The future of economics as a discipline would seem to lie in such endeavours and the contributions in this book represent clear progress in this direction.

3. OVERVIEW

The authors of the chapters in this book take up the challenge of discovering an empirically based evolutionary economics based on complex system theory. The first set of contributions deals with empirical research methodologies, discussing econometric techniques and practices, as well as new methodologies. The other chapters deal with the application of evolutionary principles and ideas to economic questions.

In Chapter 2, Foster explores, in a critical way, the current conventional practice and methodologies of time-series econometrics from an evolutionary perspective. He assesses the vector error correction modelling methodology (VECM) prevalent in time-series econometrics. The main goal of the VECM is to isolate stable long-run parameters. By allowing the unknown disequilibrium dynamics to be captured, proponents of the VECM argue that their methodology is dynamic. However, as Foster remarks, structural change associated with regulatory change and innovations cannot be captured in such an approach. He points out that this methodology diminishes the extent to which time-series econometrics can be used to discover new explanations of economic phenomena that reflect the evolutionary character of the economic system. By drawing on joint work with Phillip Wild (Foster and Wild 1999), he shows that a quite general econometric methodology grounded in evolutionary economics can be specified. Foster uses the consumption function as an example and shows how evolutionary economics, understood as a set of theories concerning economic processes, is able to provide a different econometric methodology which respects economic history, captures selection processes, and reflects the self-organizational nature of the economy. The equilibrium/disequilibrium constructs of the VECM time-series econometrics are replaced by non-equilibrium visions of economic change at the macroeconomic level. Foster suggests that this provides not only a theoretical agenda for evolutionary economics but also an empirical one. Only by showing the relevance of evolutionary economics in applied work – especially at aggregate levels – can the habit of applied economists to rely on long-run equilibrium constructs be broken. The proposed empirical methodology allows economic history into macroeconomics, which becomes a body of theory concerned with systemic connections and the dynamics of structural change.

In Chapter 3, Bevilacqua and van Zon provide evidence that does not support real business cycle theory which, in a number of respects, represents the worst recent example of simplistic theorizing in economics. They present robust evidence that serially uncorrelated residuals of many economic time series contain non-linear signals. First, they apply statistical tools to artificially generated time series that have dynamics similar to a random walk. They show that there is a large information gain by modelling such data using non-linear methods. They go on to apply the same methods to a large number of US time series and show that there is still relevant information in the residuals of time series. Their findings imply that the standard econometric practice, grounded in real business cycle theory, of interpreting the data using linear models that are disturbed by exogenous random shocks is not warranted. The dynamics of macroeconomic time series appear to be intrinsically non-linear. They conclude that real business cycle theory, and unit root autoregressive models are generally inadequate devices for understanding economic time series and argue that their results provide evidence for the hypothesis that economic variables may not follow a stationary path even in the absence of external shocks. The observed non-linearity and non-stationary are what we would expect if economic systems were complex and self-organizational in character.

In Chapter 4, Ebersberger and Pyka argue that genetic programming provides a useful modelling methodology to add to the toolbox of evolutionary economics for both theoretical and empirical research. While it is well known that genetic programming is an appropriate modelling strategy to capture the learning of agents with limited computational power, Ebersberger and Pyka suggest that genetic programming can be used to improve existing models and support the generation of new economic models that can be related to empirical observations. The selection and rearrangement of rules can be modelled with genetic programming. By asking what modelling is, they show that the learning of agents needs to be modelled as a modelling process. They discuss how an economist can be modelled as a bounded rational agent in search of principles that can explain empirical regularities, and show how genetic programming was employed to make a theoretical model, concerned with the presence of twin peaks in world income distribution, more parsimonious. Regarding the applicability of genetic programming techniques to empirical analysis, Ebersberger and Pyka note that genetic programming is especially suitable in applied evolutionary economics, as it provides a non-parametric framework, can take a priori knowledge into account, allows for heterogeneity and can be used to detect structural breaks.

In Chapter 5, Frenken and Nuvolari show how the NK model of evolutionary biology can be employed in empirical studies of technical change through the use of entropy statistics. The NK model developed by Stuart Kauffman (see Kauffman 1993) is a simple, formal model that illustrates the role of systemic interdependencies in complex systems. Many scholars have recognized that interdependencies between technological components are the prime source of design complexity. The NK model can represent the design process employed in producing a complex technological artefact as a trial-and-error process that is bound to end up in a local optimum. While the NK model has received considerable attention, empirical applications are relatively scarce. Frenken and Nuvolari show how the use of entropy statistics allows a relatively straightforward application of the NK model to empirical studies of technological change. The approach enables the study of evolutionary patterns in technological change to be examined in terms of variety and differentiation. They propose a number of generalizations of the original NK model to account for the specific attributes of technological evolution. They conclude that the evolutionary development of a complex technology, following the (generalized) NK model, is expected to be characterized by both an increasing degree of variety and an increasing degree of differentiation, when the complex nature of technological artefacts and heterogeneous demand is taken in account. They provide applications of this approach to the design dimensions for three different technologies, namely aeroplane, helicopter and early steam-engine technologies. Their conjecture of increasing variety through differentiation is confirmed for aircraft and steam engines – both aircraft technology and early steam-engine technology were affected by the introduction of a revolutionary design (the jet engine and Watt's engine, respectively). However, in both industries this did not lead to a substitution process but to a process of progressive differentiation into different design families. The case of helicopter technology shows the opposite of a differentiation process. Here Frenken and Nuvolari argue that this can be attributed to the presence of competing aircraft models. Their empirical results offer an important insight. The evidence for aircraft and steam engines shows that the evolution of complex technologies is better described as an evolutionary process of differentiation than by the model of linear substitution. This is entirely in accord with complexity theory.

In Chapter 6, Reinstaller and Hölzl present an analytical framework that deals with induced innovation to study the nature of recombinant search based on the NK model. They take issue with the view that technological search is essentially a random process and outline a framework where interdependencies between production processes and inputs drive the adoption of new processes. Complementarity makes the production system non-separable and complex. Strong interdependencies between specific elements create a 'core' process technology. The 'core' can be understood as akin to a dominant design for production processes. The property of the core is that it restricts the search space for variety generation. The search for variations of the design is limited to the neighbourhood of the core. A change of elements in the core would increase the likelihood that changes in one element would affect the overall performance. Thereby, technological uncertainty is reduced. The other side of complementary relations is that they may cause imbalances and turn into binding constraints. The break-up of complementary relations increases technological uncertainty and triggers the search for new combinations. Complementarities assume the role of focusing devices. Reinstaller and Hölzl argue that strategies of problem decomposition and re-composition are applied in order to soften the constraints. This leads towards a modularization of activities. They apply their framework in a historical case study of the establishment of the first IT regime at the turn of the nineteenth century in the USA. They argue that both the accounting revolution and the adoption of office machinery were guided by complementary constraints.

In Chapter 7, Grebel, Pyka and Hanusch provide an eclectic, evolutionary model of entrepreneurship. The first part of the chapter contains an overview of the theory of entrepreneurship. From this they distil the most important aspects of entrepreneurship, which serve to locate their model in the literature. The building blocks of their model are bounded rational agents. These agents are modelled to have a specific set of attributes. Each agent is equipped with 'entrepreneurial spirit', human capital and venture capital. The 'entrepreneurial spirit' captures the entrepreneurial function emphasized in Schumpeter's theory of entrepreneurship. This feature describes the tendency to become an entrepreneur or an employee. Human capital refers to the investment in knowledge and know-how in the agent. According to their theory, human capital is the crucial productive element that decides the fate of the entrepreneurial firm once it is established. Venture capital refers to the financial means necessary to set up a new firm. New firms are set up by networks of agents. A random matching process brings the agents together and this coalition evaluates the possibility that a new firm can be formed. The survival of the newly founded firm depends on human and venture capital. The authors use a simulation study to show that firms do not enter all at once and that turbulence is greatest in states of emergence. As time goes by, it becomes more and more difficult to enter. This is captured by the entry threshold, which relies negatively on the growth rate of the sector's turnover and positively on exit. By using evidence for the set-up of new Internet/e-commerce firms in Germany, Grebel, Pyka and Hanusch are able to show that the results of their model fit the facts of new firm formation in newly emerging sectors.

In Chapter 8, Krafft offers a critical perspective on the concept of shakeout in the industry dynamics and industry life cycle literature. Many industries evolve according to a life cycle, and firms in those industries eventually face a shakeout as the opportunities for variety generation decline and a dominant design is established. Such a shakeout follows on from a process of competitive escalation in advertising and/or R&D. However, as Krafft shows, a number of industries do not evolve along a typical life cycle trajectory, either because they are essentially knowledge-driven instead of technology-driven, or because they exhibit patterns of evolution that do not involve shakeout. Krafft points out that knowledge-driven industries, in particular, often show pronounced non-shakeout patterns in their industrial dynamics. She shows that there have been no real shakeout patterns in the chemical, telecommunications and the medical instruments industries. In those industries, networks, clusters and alliances prevented a shakeout from occurring. Krafft argues that shakeouts that do occur in knowledge-driven industries have different patterns from those in technology-driven industries. This is because knowledge-driven industries have different network structures. Using the evidence from selected industries, Krafft argues for a wider research agenda in the study of industrial dynamics. To understand the main determinants of industry evolution, we cannot restrict attention to aspects of product and process technologies. Relationships between firms, suppliers and customers are vitally important in examining specific cases of industrial dynamics, as is the diversity in the institutional set-up between countries. In essence, Krafft shows that taking complexity seriously in examining industrial dynamics means that the life cycle, which is an analogy drawn from biology, is not an adequate representation of the self-organizational and interconnected dynamics of firms and industries.

In Chapter 9, Peneder deals with the so-called 'Austrian Paradox'. This involves an ambiguity between pronounced deficits in an industrial structure and a general perception of its good macroeconomic performance. Peneder reviews evidence which confirms that Austria has, in comparison with other developed countries, low shares of technologically progressive industries and low R&D levels. This negative assessment is in sharp contrast to Austria's strong macroeconomic performance, with comparatively high levels of labour productivity and income, high GDP growth and below-average unemployment rates. The Austrian experience seems to conflict with the basic evolutionary thesis that structural change drives economic growth. Peneder then provides evidence that the particular path of Austria's successful macroeconomic development can be explained by specific geographical, socio-political and institutional factors, especially (i) industrial relations, shaped by the corporatist institution of social partnership which promotes aggregate wage flexibility, (ii) close proximity to the dynamic high-income regions of southern Germany and northern Italy, (iii) the presence of anti-cyclical fiscal policy through built-in stabilizers in the generous system of social security, public investment and accelerated investment schemes and (iv) a strong currency, which has kept inflation under control. Taken together, these policies stabilized expectations and created a favourable climate for private investment. Good macroeconomic performance is related to the unconventional finding that macroeconomic stabilization has had a lasting impact on the level of GDP per capita. Peneder then goes on to show that the long-term coherence between macroeconomic policies and institutions is no longer given. The membership in the Economic and Monetary Union led to a loss in autonomy in the formulation of national macroeconomic policies. However, even before this, the growth regime had lost much of its lustre. But, given the constraints, Peneder argues, a new policy mix will have to be based primarily on the supply side; that is, industrial policy will have to be used. Peneder goes on to present an outline of an evolutionary (Schumpeterian) perspective on industrial policy which projects evolutionary dynamics into a coherent set of policies directed towards structural change and growth. This set of policies is derived from the evolutionary principles of variety generation, diffusion and selection. For Peneder, industrial policy in an evolutionary spirit should foster (i) novelty generation through entrepreneurship and innovation, (ii) the accumulation of productive resources through knowledge creation and diffusion, and (iii) open and competitive markets to provide an effective means of selection.

In Chapter 10, Tappi deals with structural change and development at the regional level. She has studied changes in industrial specialization in the Italian Marches region from the production of musical instruments towards electronics. She interprets this change as an outcome of a process of self-organization based on the differential behaviour of firms. With the exception of the presence of the University of Ancona, which guaranteed a high-quality labour force with engineering know-how, Tappi argues that the changes were orchestrated by the (unintended) heterogeneity of behaviour and strategies and also organizational forms of the single firms. Larger firms were instrumental in providing the absorptive capacity of the regional economy. However, the collective learning process, which embedded new technological knowledge into the regional competence set, was largely driven by small start-up firms. She provides evidence for the thesis that localized competence and localized collective learning processes are an important factor in cushioning the risk of entrepreneurial activity. Networking, which is central in complexity theory, is fundamental in this case and, through careful research, Tappi has demonstrated that the more conventional selection models favoured in evolutionary economics are not rich enough to capture the regional developments that she has studied.

In Chapter 11, van den Bergh examines the interrelations between ecological and evolutionary economics. He emphasizes the relevance of evolutionary thinking to the understanding of environmental problems and provides a concise overview of key issues in environmental and resource economics. He argues that ecological economics is a field that emerged in the 1980s because of concerns with the sustainability of economic development and, as multidisciplinary analysis, he sees it as close in spirit to evolutionary economics. The relationship between evolutionary economics and environmental economics is highlighted by focusing on the themes of economic growth, environmental quality and the role of resource management in major structural changes. The most fundamental task that van den Bergh identifies is the formulation of an evolutionary theory of growth which takes environmental resources and needs, as well as a spatial dimension, explicitly into account. The spatial dimension is especially important for environmental and resource problems, as they are spatially heterogeneous. This has to be considered in the formulation of any environmental policy. However, there is a problem in using evolutionary economics as a theoretical basis for the formulation of policy instruments, as van den Bergh remarks. That is, there is no normative evolutionary welfare theory that can match neoclassical welfare theory. However, policy suggestions can be derived from evolutionary economics, especially if bounded rationality, endogenous preferences, path dependency and the variety (knowledge) generation are considered to form the core of evolutionary economics. Once again, complexity theory is likely to be helpful in enriching evolutionary economics in a way that can lead to a better understanding of policy advice. Conceptualizing policy formulation is a dimension of knowledge formation and is, itself, a complex and fragile process of communication and consensus building. Van den Bergh provides what can be viewed as a first step towards more integration of economic and environmental policies built up from a more consistent set of analytical principles.

REFERENCES

Depew, D.J. and Weber, B.H. (1995), *Darwinism evolving: system dynamics and the geneology of natural selection*, Cambridge, MA: MIT Press.

Dopfer, K. (2001), 'History-friendly theories in economics: reconciling universality and context in evolutionary analysis', in J. Foster and J.S. Metcalfe (eds), *Frontiers of evolutionary economics*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Dopfer, K., Foster, J. and Potts, J. (2004), 'An evolutionary framework for economic analysis', *Journal of Evolutionary Economics* (forthcoming).
Fagerberg, J. (2003), 'Schumpeter and the revival of evolutionary economics: an appraisal of the literature', *Journal of Evolutionary Economics*, **13**, 125–59.
Foster, J. (2000), 'Competitive selection, self-organization and Joseph A. Schumpeter', *Journal of Evolutionary Economics*, **10**, 311–28.
Foster, J. (2003), 'From Simplistic to Complex Systems in Economics', paper presented at the Economics for the Future Conference, Cambridge (UK), 17–19 September.
Foster, J. and Metcalfe, J.S. (eds) (2001), *Frontiers of evolutionary economics*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Foster, J. and Wild, P. (1999), 'Detecting self-organisational change in economic processes exhibiting logistic growth', *Journal of Evolutionary Economics*, **9**, 109–33. Reprinted in U. Cantner, H. Hanusch and S. Klepper (eds), *Economic Evolution and Complexity*, New York: Physica-Verlag, 2000, pp. 159–84.
Hayek F. (1945), 'The use of knowledge in society', *American Economic Review*, **35**, 519–30.
Kauffman, S.A. (1993), *The Origins of Order: Self-Organisation and Selection in Evolution*, New York: Oxford University Press.
Metcalfe, J.S. (1998), *Evolutionary economics and creative destruction*, London: Routledge.
Metcalfe, J.S., Foster, J. and Ramlogan, R. (2003), 'Adaptive Growth', University of Manchester, Centre on Regulation and Competition Working Paper No. 59.
Mirowski, P. (1989), *More Heat Than Light: Economics as Social Physics*, New York: Cambridge University Press.
Nelson, R.R. (1995), 'Recent evolutionary theorizing about economic change', *Journal of Economic Literature*, **33**, 48–90.
Nelson, R.R. and Winter, S.G. (1982), *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press.
Potts, J. (2000), *The New Evolutionary Microeconomics*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Saviotti, P. (1996) *Technological evolution, variety and the economy*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Witt, U. (ed.) (1993), *Evolutionary Economics*, Aldershot, UK and Brookfield, USA: Edward Elgar.

2. Econometrics and evolutionary economics

John Foster*

1. INTRODUCTION

In the 1980s, my interest in evolutionary economics was stimulated by dissatisfaction with the econometric methodologies available to undertake applied work using time-series data. Many of these dissatisfactions were raised in *Evolutionary Macroeconomics* (Foster 1987). Although this led to an interest in evolutionary economics in its own right, my original research programme, dedicated to the development of an econometric methodology that could encompass both conventional and evolutionary modelling, remained intact. Over the years, it has been noticeable that many of the difficulties encountered in econometric modelling have stemmed more from the economic theory used by modellers than the statistical techniques that are applied. It gradually became apparent that, if the relationship between econometrics and economic theory could be respecified in a new way, this might lead to a shift away from the dominant neoclassical theoretical paradigm in economics towards theoretical perspectives embodied in neo-Schumpeterian evolutionary economics.

By 1987, conventional econometric methodology had split into three strands (see Pagan 1987). Two of these strands – VAR modelling, pioneered by Christopher Sims, and error (or equilibrium) correction modelling (ECM), popularized by David Hendry – were combined into a unified VECM approach in the 1990s (see Pagan 1994). Over the same period, my own 'evolutionary' approach to econometric modelling was developed, beginning with Foster (1991) and (1992) and ending with Foster and Wild (1999a, 1999b). These papers were greeted not with criticism, but with an uneasy silence amongst conventional econometric modellers. Conversations with such modellers have revealed that there are, in fact, no strong technical objections to the methodology proposed but, rather, an unwillingness to accept the evolutionary approach to economic theory that is adopted.

Furthermore, the Foster and Wild papers do not deal with macroeconomic times-series data but rather a specific example of a developmental process following regulatory changes. Although Foster (1992) did deal with an issue that had attracted considerable attention in the 1970s and the 1980s (the discovery of a stable econometric model of the M3 money magnitude), the methodology used was still in its early stage of development and was viewed by many modellers as, in effect, a modified ECM approach. The rise to dominance of the VECM methodology in the 1990s made it increasingly difficult to promote alternative approaches. This being said, econometric modellers are quite open to revisions in economic theory since they mostly hold the view that the development of economic theory is not their business – economic theory is something to be taken as given. None the less, there is a strong preference for economic theory of the neoclassical equilibrium kind, due to the fact that almost all econometricians have received training in it in the past or because it is so widely accepted in the literature that they read. The challenge is to persuade econometricians that evolutionary economics is capable of offering analytical principles that can be operationalized in econometric models in a manner that is more congruent with the historical nature of the time-series data used.

On the other side, evolutionary economists have had a tendency to reject conventional econometric methodology for very understandable reasons, and many have moved on to, for example, simulation approaches to modelling non-linear systems. However, evolutionary economists should reconsider the role of econometrics in their research programmes since, from a scientific standpoint, it is vitally important that theoretical models relate to the data we have at our disposal, whether they are the product of experiments and sample surveys or the efforts of statisticians to construct historical data series. In the words of Joseph Schumpeter:

> The only way to a position in which our science might give positive advice on a large scale to politicians and business men, leads through quantitative work. For as long as we are unable to put our arguments into figures, the voice of our science, although occasionally it may help to dispel gross errors, it will never be heard by practical men. They are, by instinct, econometricians all of them . . . (Schumpeter 1933, p. 12)

In Section 2, I shall discuss the difficulties with the predominant approach to econometric modelling. In Section 3, using the example of 'endogenous' growth theory, I shall explain how this methodology cannot adequately address the needs of economists. In Section 4, I summarize the Foster and Wild (1999a, 1999b) evolutionary econometric methodology that can be applied in developmental phases of economic evolution. Section 5 contains some discussion of how this approach could be generalized to deal with aggregated time-series data – the case of aggregate consumption expenditure is discussed. Section 6 contains some concluding remarks.

2. ERROR CORRECTION MODELLING: TRADING EXPLANATION FOR FORECASTS

The application of econometrics to time-series data rose to prominence in applied economics following on from the 'great debate' between Jan Tinbergen and Maynard Keynes in the 1930s (see Morgan 1989 and Dopfer 1988 for contrasting positions). Despite the warnings of Keynes concerning the limitations of econometrics, both as the basis of model building and as a medium for subjecting theories to test, it became a central tool in the application of Keynesian economic principles in the post-war era as economists tried to establish their scientific credentials through increased measurement and quantification. Econometrics became used for two purposes in the domain of time-series data: first, to test hypotheses in the absence of laboratory experimentation, and second, to provide forecasts to aid policy-making.

Considerable difficulties arose with both of these research programmes. This gave rise to much debate in the late 1970s, following the proliferation of econometric studies that drew unwarranted conclusions concerning the validity of hypotheses derived from economic theory. Attempts to provide robust forecasting models for policy-makers were also thwarted in the wake of a series of spectacular breakdowns in previously estimated econometric relationships – the Phillips curve and the demand for money function are two well-known examples.

By 1980, econometric modelling was in a state of crisis and David Hendry (1980), Christopher Sims (1980) and Edward Leamer (1983), in their own distinctive ways, offered what they saw as improved methodologies for undertaking econometric research (see Pagan 1987 for an assessment). Twenty tears later, after many twists and turns, two of these methodologies became integrated and extended to form a widely accepted econometric methodology. The prevailing vector error correction modelling (VECM) methodology contains the following features:

- Spurious regression, due to common time trending, is largely avoided
- Hypotheses drawn from economic theory tend not to be tested directly, but instead, enter time-series econometric models as 'equilibrium' restrictions
- 'Disequilibrium' processes are viewed as theoretically intractable so these are modelled statistically through a process of induction from the 'data generating process' (DGP)
- Vector autoregressive specifications are used to capture the interactive connections between data series
- A large battery of diagnostic tests, including out-of-sample forecasting and parameter stability tests, are employed routinely.

Hendry (1995) provides a *magnum opus* in which this methodology is spelt out in more detail.

What is most striking about this methodology is the marked shift that has occurred away from hypothesis testing, which has tended to become the focus of cross-section econometrics and experimental economics, towards the priorities of forecasters and policy-makers. The main goal is to isolate estimated 'long-run' or 'equilibrium' parameters that are stable enough to be useful for policy-making. Any models that do not display these features are set aside, not so much because they lack validity or explanatory content but because they are not useful in this sense.

Thus the methodological stance is very instrumentalist in orientation. However, the relegation of economic theory to only a supporting role, such as in the selection of long-run equilibrium restrictions in models that are primarily statistical in construction, has led to a severe slowdown in the development of new theoretical perspectives, particularly in macroeconomics. 'Good' theory must have a strong logical base that can provide the foundation for quantifiable predictions which can be used to generate equilibrium state predictions. Neoclassical economics is compelling in this regard and almost universally preferred. However, neoclassical economics is very much concerned with individual optimization, so very strong aggregation assumptions are necessary to connect this kind of economics with the macroeconomic level of enquiry. This usually results in the adoption of the fiction of a one-agent economy (the 'uniform' agent is often referred to as a 'representative' agent, but this is clearly misleading).

'Classical' hypothesis testing, involving hypotheses derived from economic theory, becomes non-viable, as the abstract and timeless character of economic theory necessitates a 'translating' hypothesis (typically a partial adjustment or error learning hypothesis in the pre-ECM days), which must be introduced to allow the main hypothesis to address historical data. This approach is bedevilled by the fact that the econometric specification tested is a combination of two hypotheses resulting in uncertainty as to what is being tested, and the recurrence of observational equivalence problems. In the ECM or VECM approaches, there is no attempt to specify any particular translating hypothesis. Instead an inductive statistical analysis of the DGP, which is presumed to capture 'disequilibrium', is conducted. The 'best' statistical representation of the DGP is chosen, in conjunction with a theoretically based hypothesis, which is represented by an 'error (or equilibrium) correction term' (ECT) and appropriate parametric restrictions, provided that co-integration tests are satisfied. Equation (2.1) contains the typical set-up:

$$[\ln Y_t - \ln Y_{t\text{-}1} = a + b(\ln X_t - \ln X_{t\text{-}1}) \ldots DGP \text{ etc.} \ldots - c(ECT) + u] \quad (2.1)$$

Of course, some hypotheses will perform better than others as restrictions – this is explored through 'nesting' and 'encompassing' tests but, ultimately, whatever hypothesis is chosen cannot be viewed as being conclusively supported because of the unknown nature of the DGP. Instead, what can eventuate is a model with good in-sample forecasting and parametric stability properties. So what we have is not an explanatory methodology, but rather one concerned with obtaining statistical representations of history that are coherent and appear to bear some relation to economic logic. Models that perform well but have 'perverse' estimated coefficients tend to be rejected.

Research dealing with the 'demand for money' provides a good example of how econometric modelling has been conducted. From the early 1970s onwards, forecasting and parameter stability failures were encountered in estimating such models. These models contained, with minor variations, a monetary aggregate, an interest rate, an income measure and a price index, in line with prevailing demand-for-money theory. However, such failures were not met with a fundamental rethinking of the theory involved, but rather with some 'band-aid' theoretical modifications and a myriad of suggested statistical solutions. By the time Foster (1992) and others espousing variants of the 'endogenous money' approach offered econometrically supported explanations of the breakdown that *did* involve a major rethinking of the theory being used to understand how money magnitudes are determined, there was little interest simply because the findings involved 'perverse' or 'heterodox' theorizing.

Furthermore, the tendency to espouse an instrumentalist, rather than an explanatory, methodology led to the conclusion that these new developments were of little 'use' since most governments and their central banks had, by that time, ceased in their attempts to control money magnitudes and, in any event, evidence supporting hypotheses that money is endogenous implied that monetary control is extremely difficult. Significantly, the flow of macroeconomic theories embodying a strong presumption concerning the existence of a stable and well-determined demand-for-money hypothesis (particularly in new classical macroeconomic theory) continued unabated despite the absence of convincing supporting evidence and the availability of alternative theoretical perspectives that received stronger econometric support.

Thus what occurred was clearly a very different business to 'normal science', whereby the refutation of a theory leads to its abandonment. At one extreme, there emerged theories that bore no connection with reality whatsoever. For example, real business cycle theory is a logical construction that makes no contact with the reality of history beyond borrowing econometrically estimated parameters from very diverse and inconsistent sources for calibration exercises. At the other extreme, the ECM modellers have sometimes used empirically unsupported hypotheses, drawn from economic theory, in setting 'long-run' equilibrium restrictions, on the grounds that the lack of support was really due to the problem of the unknown qualities of the DGP. The only 'reality check' in this approach involves tests to see if the variables in question are co-integrated but, of course, this says little about the presence of theoretical equilibrium relationships and, in any event, is no more than a test of the relationship between the 'average disequilibrium' states of variables if such an equilibrium/disequilibrium dichotomy is employed.

The strength of the ECM methodology lies in its proven capacity to deliver models that provide good forecasting performance, at least over short periods. Paradoxically, however, policy-makers have come to rely less on econometric models, simply because macroeconomic policy has become much less activist than it was in the eras of monetarism and Keynesianism. Yet without the employment of models by policy-makers, instrumentalist methodologies become rather pointless.

In the face of this, ECM proponents have attempted to make more of their methodology than is wholly warranted. In particular, it is claimed that the methodology is more 'dynamic' than others. What this seems to mean is that, by capturing the statistical features of the DGP well, unknown 'disequilibrium' dynamics are dealt with, leaving a clearer picture of 'long-run equilibrium' relationships. This is reflected in the title of David Hendry's (1995) *Dynamic Econometrics*. However, on close inspection, there is little that is dynamic in its pages. The theories from which 'long-run equilibrium' hypotheses are drawn are predominantly static and timeless and there is little or no discussion of the structural change that is manifest in all historical processes that are contained in time-series data.

Significantly, the only structural change addressed in any depth by Hendry (1995) is the 'Lucas critique', which relates to the impact of forward-looking expectations on the parametric structure of a macromodel. However, the goal is not to provide a new way of dealing with structural change but rather to find ways of continuing to use fixed parameter methods through the introduction of ideas such as 'super-exogeneity'. In fairness, Hendry (2000) does explore the issue of structural change in more depth and he concludes that the conventional approach is poor at detecting structural change in the DGP except when there is a 'mean shift' in the 'equilibrium' relationship. However, there is little or no discussion of the theoretical aspects of structural change processes. Explanatory content is not a primary concern but, as always, forecasting performance is.

One solution to the problem of structural change is to adopt a varying parameter method, such as the Kalman Filter, but this method does not even get a mention in the 869 pages of Hendry's (1995) book. A possible reason for this is that such an approach results in a significantly reduced ability to undertake systematic tests of significance, which are 'core' in the ECM methodology. Thus there is no basis for achieving such things as 'parsimony' or offering stable estimated parameters to policy-makers. Again, this has been recognized elsewhere by David Hendry, who acknowledges the difficulties presented by 'inherent non-stationarity due to innovative human behaviour or natural processes' (Doornik and Hendry 1994, p. 295).

This difficulty tends to be understated and viewed as something that might be 'meta-parameterized', but few examples are provided. Clearly, this problem is likely to be very important in modelling variables such as aggregate business investment, and it is indeed the case that the ECM methodology does not perform well by its own standards in this context. It is in coming to an understanding of how structural change operates that we truly engage with 'dynamics' in the economic system. And it is precisely because policy-makers intuitively understand that structural change – associated with regulatory changes and rapid innovation – shifts parameters that they are wary of fixed parameter econometric modelling built upon a DGP of unknown pedigree.

3. THE ECONOMETRICS OF 'ENDOGENOUS' ECONOMIC GROWTH

In examining the recent literature in applied econometrics, a paradox becomes evident. Error correction modelling is the pre-eminent approach in applied econometrics but, in the expanding literature concerning 'endogenous' growth theory, attention has centred upon the fact that there exists a significant constant term (TFP growth) in estimated growth equations, relating the growth of output to the growth of capital, labour and other time-differenced variables. The existence of such a constant implies that the variables of interest do not move together – they diverge over time and are not co-integrated. In such circumstances, the ECM methodology advises that a first-differenced ECM model should not be estimated, but that the residuals of such a growth equation could be checked to see if they are I(0). If so, an ECM can be constructed using second-differenced data. But proceeding in this manner completely misses the point because, in distracting attention away from the non-stationary dimensions of the growth process, the central *theoretical* focus of endogenous growth theory is also eliminated.

Although endogenous growth theorists tend to remain in the neoclassical economic paradigm, they are not content simply to seek statistical configurations to address the empirical inadequacy of neoclassical growth theory. Unlike the demand for money, to which modellers alluded to earlier, their goal has been to *extend economic theory* to better understand the observed non-stationary aspects of economic growth. To this end, they discuss the nature of technical change, creativity, innovation, learning-by-doing, knowledge externalities and so forth, and they attempt to give these expression by extending the neoclassical growth model. Historical tendencies in time-series data are therefore not viewed as just the product of a data generating process that involves intractable 'disequilibrium', which requires statistical procedures to deal with it. There is a genuine attempt to find new theoretical explanations. Thus it seems that those who deal with the econometrics of economic growth have come to occupy quite a different methodological stance to conventional VECM applied by econometricians.

Although new growth theory retains the Cobb–Douglas production function set-up that lies at the core of neoclassical economics, 'dynamics' are not something associated with 'disequilibrium'. Instead, additional theoretical propositions are made, which, in effect, keep equilibrium on the move. This is also how real business cycle theorists view their models, but the crucial difference is that they are highly conservative in their application of neoclassical economics, whereas new growth theorists make no strong presumptions concerning the universality of optimization and are uninhibited in relaxing key assumptions. In their world, technological change does not emanate from exogenous random shocks but rather is the endogenous outcome of creative, learning and knowledge spillover processes that cannot be encapsulated in a traditional neoclassical general equilibrium model.

For a VECM modeller, a growth equation with a significant constant is unacceptable because it does not have 'sensible' long-run equilibrium properties. In contrast, new growth theorists use equilibrium theory as a starting point and then argue that it must be modified in important ways to address the data properly. A key distinction between the two approaches is that VECM modellers derive their first-differenced specifications from atheoretical representations of the data involving combinations of lagged and lagged-dependent variables, plus an error correction term, provided that the hypothesis that co-integration exists is supported. In contrast, new growth economists hold with economic theory in their initial specifications and derive growth equations by first-differencing the logarithmic form of an aggregate production function. However, in so doing, they accept that there is likely to be parameter variation in both the scale variable and the exponents in the Cobb–Douglas production function.

This being said, endogenous growth theorists have tended to shy away from time-series data – the bulk of new growth econometrics is not conducted on time series but instead on cross-section data across countries, using, for example, the Summers and Heston (PENN) database. Each country is implicitly assumed to be in equilibrium at whatever level of development it happens to be. From many perspectives this is unsatisfactory, given the common Cobb–Douglas production function imposed across countries as disparate as the USA and the Sudan. This is recognized to some extent by grouping countries according to additional theoretical criteria that are used to attempt to eliminate variations in TFP (see Jones 1998 for a review). However, it is not clear how the differential growth experience of *ad hoc* groups of developing countries helps us to understand the actual growth process in particular countries over selected time periods (Durlauf 2001).

The endogenous growth theorists eschew the opportunity to expose their theories to historical data and this is partly due to the fact that they take a very circuitous route in incorporating new theoretical factors that are viewed as fundamental drivers in the growth process. There seems to be a concern with preserving neoclassical theoretical foundations at the core of economic analysis in introducing radical theoretical additions. Despite this conservatism, some new growth theorists go as far as labelling their models as 'Schumpeterian', as for example Aghion and Howitt (1998). There is indeed some connection, because many of the proposed theoretical extensions have been discussed in evolutionary economics for a number of years (Nelson 1997). However, following on from Nelson and Winter (1982), evolutionary economists have appreciated that, if creativity, innovation, entrepreneurship and learning, among others, are the mainsprings of economic growth, they should be placed at the centre of any analysis of the growth process.

The manner in which new growth theorists employ econometrics inappropriately because of a fundamental confusion between 'growth' and 'development' has led many evolutionary economists to become even more strident in their rejection of conventional econometric methods. Econometrics, if used at all, tends to play a supplementary role (see, for example, Fagerberg 2000). Although such a stance is entirely reasonable, it remains true that, as long as the public looks on economics as a science, it will demand that the propositions made have some analytical credibility and can be demonstrated to be relevant to the real world, as seen through the historical data that we have at our disposal. This has to be demonstrated by evolutionary economists, otherwise they will have little impact.

One solution is to avoid econometrics and adopt the investigative techniques of the historian, as well as a compatible methodology, such as critical realism (Lawson 1997). This has been done in many interesting non-quantitative studies of, for example, innovation processes and the rise and fall of firms. However, this kind of research is often not regarded as scientific either by the public or by mainstream economists, because the sheer complexity of historical interactions allows too much play for subjective interpretations. Science is widely perceived as the isolation of simple operational principles which can be demonstrated to hold in 'average' circumstances. Because such principles are simple and incomplete, they are open to rejection when confronted by data. However, this can also mean that more than one theory can be supported by the data. Since observational equivalence is always with us, the 'reasonableness' of the analytical principles used in empirical work is very important. It is the apparent reasonableness of many of the fundamental principles embodied in neoclassical economics that has allowed it to be so powerful for so long, irrespective of the quality of the supporting evidence.

As noted, econometric evidence is not decisive in the rejection of economic theory once it has been discovered that derived hypotheses are not supported. Theoretical principles survive in economics if they are deemed to be reasonable. This means that econometric methodology becomes inseparable from broader methodological considerations concerning how we undertake economic analysis. VECM methodology is welded to ontological presumptions concerning equilibrium and disequilibrium in models with neoclassical underpinnings. How endogenous growth theorists choose to do econometrics across countries is inseparable from the neoclassical representations of economic behaviour that form the core of their analysis. Similarly, many real business cycle theorists choose calibration rather than estimation because of their beliefs about economic theory.

Economists have drawn heavily on the natural sciences in their quest to be seen as 'scientists'. However, the tendency to ignore evidence that rejects hypotheses of importance suggests that economists do not have confidence in the empirical methods that they use. They rely more upon their beliefs concerning the reasonableness of their theoretical foundations, derived as much from introspection as observation. So how can we approach the historical record in a more 'scientific' manner, one which can provide explanations we have not already decided upon due to our adherence to belief systems and ideologies? In the social sciences this is not an easy question to answer. But a shift away from a perspective where history is something that must be reconciled with timeless theoretical principles to one where theoretical principles are time contingent and must be reconciled with recorded history, is a beginning. Such a beginning has been made in evolutionary economics.

4. EVOLUTIONARY ECONOMETRIC MODELLING

Neo-Schumpeterian evolutionary economics has largely been preoccupied with the operation of selection mechanisms. Variety in productive capabilities, which stems from the existence of novelty, results in the differential growth of firms due to productivity and quality differences. If the link between profits and growth is sufficiently uniform across firms, firms with the highest profits come to dominate. Thus, growth stems from non-average behaviours and the outcome that eventuates over time is not an average of these behaviours. At first sight, this poses a difficulty in applying econometrics because it is a method that deals with average associations between data series. However, this is something of a misunderstanding because the non-average behaviour that gives rise to economic growth relates to successive points in historical time. Average associations can still exist between variables over time that are meaningful and important because of the bonding that exists in economic structures and its tendency to persist from one point in time to another.

The evolutionary mechanism provides us with a coherent account as to how historical processes, and associated economic structures, develop endogenously over time. The DGP is therefore not an intractable 'disequilibrium' precipitated by exogenous shocks, but, rather, an outcome of systematic, non-equilibrium processes that can be analysed and understood. The resolution of novelty into innovative diffusion is to a large degree endogenous, simply because capabilities emerge from novelty in response to perceived problems and opportunities in the existing environment. However, such a process need not occur smoothly and there is no doubt that exogenous shocks will also have a role to play.

Consider, for example, the case of a set of firms applying different techniques to produce a good that is sold. Initially sales will begin to grow and, as sales reach market saturation, a few firms with the best techniques will dominate because of their lower costs and/or higher-quality products. Sales data for the product are likely to trace out a logistic curve. This can be fitted as a simple historical relationship where the relationship between $X(t)$ and $X(t-1)$ exhibits a varying parameter that tends from zero to a maximum and then back to zero. The advantage of the logistic curve is that we can obtain a fixed parametric representation of such a sigmoid growth path. However, as such, it is no more than a summary of the history of a growth indicator that reflects an underlying diffusion/selection process. Thus such a curve does not depict the process but, rather, an aggregate measure of the output of the process.

Many economists have viewed such logistic curves as tracing out disequilibrium dynamics (see Dixon 1994 for a review) and/or learning trajectories (see Baba et al. 1992 for an example) that conclude in equilibrium. Curve fitting can become quite sophisticated – a recent example is Bewley and Griffiths (1999), who introduce a Bayesian approach to modelling the logistic curve. The disequilibrium approach, of course, presupposes the existence of a limit that is an equilibrium position. This contrasts with the self-organization perspective where an end stationary state is viewed as structurally unstable (see Allen 2001). In this regard, Sarkar (1998) provides an insightful comparison of the equilibrium and non-equilibrium approaches to the modelling of diffusion. However, given that the logistic equation is mathematically deterministic, the question arises as to whether it is valid to use it in non-equilibrium settings that exhibit structural discontinuities. This issue is tackled in Foster and Wild (1999a, 1999b).

Foster and Wild explain why the logistic diffusion equation can be viewed as an abstraction derived from an endogenous 'theory of historical process'. The self-organization approach to system dynamics, which views 'dissipative structures' as capable of structural development through parallel increases in order and complexity towards a capacity limit, is used as the basis for such theorizing. An augmented logistic diffusion model (ALDM), which allows the diffusion rate and the capacity limit to be subject to exogenous and interactive effects, is developed.

An ALDM can be based on several alternative logistic equations. The Mansfield variant, expressed in terms of a growth rate, was chosen because of its convenient properties:

$$\begin{aligned}\ln X_t - \ln X_{t-1} = b_1[1 - \{X_{t-1}/K(...)\} - a(...)] + b_2(...) + \\ c\{\ln X_t - \ln X_{t-1}\}_{t-1} + e_t \qquad (2.2)\end{aligned}$$

where b_1 is the underlying density dependence, or diffusion, coefficient (after allowing for deterioration rates, death rates, etc); $K(...)$ represents a carrying capacity which can vary because of exogenous external factors; $a(...)$ contains competitive factors due to the presence of other systems in the same 'niche', altering the effective capacity limit that can be attained; and $b_2(...)$ contains exogenous influences which cause the net diffusion rate to vary. The lagged-dependent variable is included to capture momentum effects which cushion the impact of exogenous shocks.

Equation (2.2) is an endogenous growth specification which can be applied in historical episodes when structural development is taking place. As it stands, many might interpret it as the specification of a disequilibrium process, following a jump in a 'long-run equilibrium' K, given that ongoing structural change is homogenized into a growth measure. However, in the presence of structural change which is self-organizational in character we cannot accept this interpretation because dissipative structures which develop structurally and thus grow towards a capacity limit are not in disequilibrium. If we rely upon self-organization theory, we can predict that, as the growth of such systems tends towards zero, they do not approach a stable equilibrium but, rather, a state of structural instability.

In general, processes involving endogenous structural change are not deterministic and cannot tend, asymptotically, to stable long-run equilibrium outcomes. However, this does not mean that the conventional notion of equilibration is inapplicable. A self-organizational process can still be viewed as a moving temporary equilibrium, which tends, asymptotically, to K. Homoeostatic mechanisms of varying strength will operate to return a process to its logistic path when external shocks are experienced. If a growth process is perceived in this way, then structural instability relates to the extent to which the basin of attraction around such moving equilibria changes over time. If the logistic growth path is viewed as capturing the deterministic component of the process of structural development, then variation in the basin of attraction can be seen as reflecting its non-deterministic component. Self-organization theory suggests that there will be a tendency for the basin to narrow as the system in question moves up the logistic growth curve. The consequent fall in variance is associated with an increase in the likelihood that a given exogenous shock will induce a departure from the basin and structural discontinuity of some type.

Typically, students of technological diffusion draw families of logistic curves over time with gaps or overlaps between them, stressing the uniqueness of each diffusion process with the gaps confirming the existence of structural discontinuities. Thus in such studies it is implicitly accepted that a tendency towards saturation in a technological diffusion process is not a tendency towards a stable equilibrium. Foster and Wild (1999a) show that the particular ALDM growth trajectory that they studied is not mean reverting. But it is also not a random walk, with or without drift. The level of the variable under investigation and the (moving) capacity limit are not linearly co-integrated. Thus, in the case considered, it is difficult to argue that the observed ALDM provides evidence in support of an equilibrium/disequilibrium process, from the standpoint of either deterministic or stochastic trends in time-series data. Foster and Wild (1999b) go on to show how spectral methods can be employed to detect the presence of self-organizational change, as well as to test for the presence of a disequilibrium process of the traditional type.

If we think of fitting simple logistic curves as ‘first-level’ econometrics in the presence of self-organizational change and the Foster/Wild approach as a ‘second-level’ approach which can capture the growth dynamics of developmental processes well beyond simple cases of innovation diffusion, there remains a ‘third level’ where we are confronted with data that are aggregated across many products and sectors in a range of evolutionary phases. This third level is crucial in linking the principles of evolutionary economics to the aggregate contexts that are so familiar in mainstream econometric modelling. Let us now consider how such links can be forged.

5. EVOLUTIONARY ECONOMETRICS IN MACROECONOMIC SETTINGS

Not only do the results in Foster and Wild (1999b) raise fundamental questions concerning the interpretation of evidence in ECM studies; they also open up the possibility that we can apply evolutionary thinking in empirical contexts where the data do not involve a visible logistic diffusion path. If we think of any economic system as a dissipative structure that imports energy and materials and export products, this structure tends to become both more ordered and more complex as it develops. However, structure by necessity involves orderings which exhibit a degree of irreversibility. This irreversibility places a boundary on the extent to which the structure can develop. We can give expression to this tendency in quite a simple way if we translate everything into value terms. We have the following flow identity:

$$Y_t = Y_{t-1} + Z_t - W_t \tag{2.3}$$

or

$$Y_t - Y_{t-1} = Z_t - W_t, \tag{2.4}$$

where Y is the value flow of output characteristics of a system, W is the output value flow loss in a system due to wear and tear, breakdowns and so forth, and Z is the output value flow increase due to new investments. Clearly, if Z exceeds W there is growth and vice versa. Part of Z offsets W and part of it represents new value creation from the production of greater output of existing products or the output of new products. This is often thought of in terms of 'replacement' and 'net' components of investment expenditure, but as Scott (1989) stressed, this can be misleading because 'replacement' often involves the simultaneous upgrading of productive structure and output. Self-organization theory does not rely on such a distinction but, rather, predicts that growth will run out as Z becomes increasingly committed to dealing with W. In other words, Y will follow a logistic curve.

This implies that Equation (2.4) must take the following form:

$$\text{If:} \quad Z = z\ Y_{t-1} - mY_{t-1}{}^2 + v \tag{2.5}$$

$$\text{And:} \quad W = w\ Y_{t-1} + n\ Y_{t-1}{}^2 + u, \tag{2.6}$$

where u and v represent non-deterministic factors.

Therefore:

$$Y_t - Y_{t-1} = (z-w)\ Y_{t-1} - (m+n)\ Y_{t-1}{}^2 + u - v \tag{2.7}$$

If $K = (z-w)/(m+n)$ and $(z-w)) = b$, then:

$$Y_t - Y_{t-1} = b\ Y_{t-1}\ [1 - (\ Y_{t-1}/K)] + (u-v) \tag{2.8}$$

The logistic curve can capture growth in the presence of self-organizational development and is quite general in the sense that even if *m* or *n* (but not both) is equal to zero, there will still be a logistic form. It is very likely that *n* will be positive, since repair and maintenance requirements in dissipative structures cause *W* to rise non-linearly because of the irreversible character of parts of the productive structure. New output change, *Z*, is likely to have a density-dependent relation with output, generally discussed in terms of economies of scale but, if there are learning effects that run out, then *m* will be positive. In this regard, we must think of *Z* as being related both to all past investment and to learning effects. The non-deterministic elements *u* and *v* are important because it is a *u* shock that starts a developmental process and a *v* shock that induces a structural transition when growth has ceased. Also, these may be externally or internally generated and may or may not be random in character.

Because of the history of logistic curves in population ecology, *K* is commonly thought of as an environmental (or niche) delimiter. However, as has been pointed out, from a self-organizational perspective it is more profitable to think of it as being to a significant degree endogenously determined by the irreversible nature of productive structure. This irreversibility results in a lack of adaptability in a process or product range as development proceeds. Indeed, although we can think of an entrepreneur imagining a market niche for a new product, it is only in the process of pursuing this goal, through the organization of people and capital, that *K* becomes defined. Often, it is discovered that no niche exists, in the sense that no positive profits accrue, and the organization collapses.

The fact that *K*s are imagined constructs and endogenously determined is what distinguishes economic self-organization from its counterparts in biology and chemistry. Although much of *K* will be determined by 'founding structure' (Hannan and Freeman 1989) there is always some scope for further endogenous shifts in it over time. So in reality, the *K* limit can involve both exogenous 'environmental' limits and endogenously determined limits. In a sense these are related, because lack of adaptability constrains the environmental range a system can enter. The non-existent, fully reversible system proposed in neoclassical economics can simply reconfigure itself to suit all environments, but we know that this breaches some fundamental laws of systems (see Foster and Wild, 1996). Without a degree of irreversibility, that is, order, there will be no dissipative structure.

If explanatory variables can be discovered that capture the role of exogenous factors in affecting b and K, then the underlying endogenous dynamics of the growth/development process can be observed. It is a remarkable fact that, particularly in innovation studies, we observe so many logistic paths that conform to Equation (2.8), even without such modelling exercises. This suggests that there is an inherent tendency for systems to exhibit behaviour roughly coincident with Equations (2.5) and (2.6). As we move from the traditional logistic equation set-up to thinking in terms of logistic processes with endogenously determined K limits plus exogenous effects on b and K, we need not actually observe a logistic curve in the data.

The danger, particularly with aggregated data, is that modellers imagine that the seemingly linear character of the raw time-series data implies that a linear equilibrium/disequilibrium methodology can be applied. To illustrate this point, it is worth considering the favourite case in conventional ECM literature, namely, the 'consumption function'. Aggregate consumption has been subject to a considerable amount of econometric research, with a range of hypotheses, derived from neoclassical 'microfoundations', considered in the literature. However, much of this research is misguided, because of the fact that C must be *systematically* related to Y in an aggregate sense even if the association between them at the microeconomic level is entirely random around an average propensity to spend. We know that people spend from income and that spending in turn becomes income. Aggregate consumption relates to the whole economic system and all components of expenditure eventually filter back into consumption expenditure, as Keynes so clearly explained. When Milton Friedman specified his so-called 'permanent income hypothesis' as a relation between consumption and lagged values of current income, and produced a cumulative marginal propensity to consume close to unity, he did no more than confirm a simple Keynesian systemic truth concerning the operation of the multiplier process over time.

The evolution of consumption expenditure over time is a strongly endogenous process with feedback – growth in aggregate consumption expenditure occurs, in terms of Equation (2.4), through the emergence of novel consumption (Z) in excess of abandoned expenditure on obsolete and unfashionable products (W). Such growth requires lively entrepreneurship, both to produce novel products and to produce existing products more efficiently, as well as competitive processes which can eliminate old products and their producers effectively. Per capita growth occurs because more value (and income) is created than is destroyed. But this is not a process without limit, and waves of economic activity are generated at different time scales because of the way that institutional, organizational and technological factors are utilized. Of course, exogenous factors will also play their part, with the systematic affecting both b and K and the unsystematic affecting u and v. For example, the sophistication of market, contract and consumer credit mechanisms is likely to be important for the former parameters, while governmental shifts in monetary and fiscal policy are likely to change u and v in an exogenous manner, inducing 'jumps', 'kick-starts' and 'breakdowns'.

The feedback of consumption on itself, via income and expenditure, means that aggregate income is not a simple limiting factor on aggregate consumption, *ceteris paribus*. Despite this, as being the product of a dissipative structure of interrelated preferences and products, consumption must conform to Equation (2.3). Each period, there will be an amount of 'systemic' consumption, which carries over after a drain of unrenewed consumption (W) is deducted from Y_{t-1}, because some goods and services lose favour with customers. There will also be a flow of new consumption, as firms produce novel products and services and market them (Z). Underlying this 'consumption churning' is a myriad of product cycles, with products at different points on them. How these add up in value terms depends upon the proportions of emergent and mature products consumed. If the former dominates the latter, we observe higher per capita consumption growth.

This is an evolutionary story with the emergence of novelty, both in the products sold and in the organization and techniques used to provide them, and selection effects. The latter act in product space, not only between similar products but also across dissimilar products because of economy-wide budget and credit constraints. Some products will become 'systemic', while others are born and others die, often along with the firms that produce and sell them. The systemic circularity between consumption and income depends on a range of organizational and institutional connections (Potts 2000). Such connections also allow for the turnover of products and processes that are so crucial to economic evolution.

The macroeconomic feedback mechanism that links consumption and income has been discussed. To a Keynesian, this is familiar. However, to someone interested in fluctuations in the economy, the issue is why per capita consumption (and income) varies so much. A Keynesian would tell us that it is due to a lack of 'effective demand', which is accurate but begs the question. New Keynesians would attribute it to 'market failure' of some type, which usually turns out to be some kind of 'just so' story about coordination problems. From the self-organization perspective, consumption per capita has a K limit, which is determined by the prevalence of institutional/organizational/technological factors that are often discussed in the literature on Kondratiev waves. Once again, it is the irreversible aspect of these facilitating factors that is a key determinant of such a long-term K limit, in addition to exogenous forces.

Shorter-term business cycles occur because firms are prone to cluster their investment and disinvestment decisions, causing oscillations in per capita consumption akin to those described in the multiplier accelerator theory propounded by Paul Samuelson except that each expansion and contraction will have a different parametric structure and, therefore, no tendency to equilibrium. There will be no hard Hicksian limit cycle but, rather, a soft logistic growth form giving way to a sharper path of decline. The analysis of such fluctuation always comes back to the irreversibility of structure, its impact on costs and profits, and the fact that it is very difficult to discern outcomes in a complex system, encouraging imitative behaviour with reinforcing effects. If logistic paths for products are evenly spread in time, rather than clustered, then there should be little sign of aggregate 'waves' in per capita consumption.

Because income and consumption are systemically connected, it is no surprise to discover that they are 'co-integrated'. Neither is it any surprise that they 'error correct' over time, given their systemic association. Equally, it is no surprise to find that both per capita consumption and income follow 'stochastic trends', given that they are to a large degree endogenously determined. Surges in consumption are driven by the contagious emergence of novelty in products and processes. Reversals occur because of a contagious withdrawal from business investment in the face of profit declines arising due to over-commitment. This is not a 'disequilibrium' process, but one of creative destruction that is fundamental to the functioning of capitalism.

In depicting movements in macroeconomic time series in this way, a theoretical perspective is introduced which differs sharply from the neoclassical uniform agent approach. Instead of placing an optimization problem at the heart of macroeconomic analysis, systemic connections and the dynamics of structural change become the central concerns. Equilibrium/disequilibrium constructs are replaced by non-equilibrium visions of economic change at the macroeconomic level. This alters the manner in which we conduct econometric modelling exercises and how we interpret the results of our efforts. The connection between economics and econometrics is altered in a way that renders neoclassical general equilibrium perspectives untenable because they cannot be made congruent with the reality of structural change or the historicity of time-series data.

6. CONCLUSION

In this chapter, it has been argued that conventional econometric methodology – error correction models based on co-integrated variables and cast in VAR representations of the data – are essentially forecasting strategies that contribute little to explanation. As the 'endogenous' growth literature demonstrates, this methodology removes precisely what is of most interest in seeking new theories of economic growth, an understanding of why we observe non-stationary behaviour in macroeconomic time series. It is generally the case that conventional methodology sacrifices theoretical development in favour of forecasting priorities. It is clear that the near-abandonment of the time-series data domain, as a place for discovering new explanations of economic phenomena, creates an opening for neo-Schumpeterian evolutionary economists.

Neo-Schumpeterian evolutionary economics has the advantage that the 'theories of process' that are discussed are congruent with historical data. Self-organization and selection processes are diffusional and cumulative in nature and, as such, can be mapped directly on to historical data. The Foster and Wild (1999a and 1999b) papers have been reviewed to demonstrate that there exists an effective econometric methodology which moves beyond the large literature of logistic curve fitting into a general framework for modelling all time series. The example of aggregate consumption was discussed in this regard, simply because it is the aggregate variable that conventional modellers claim to be able to model most effectively. It is argued that the approach they adopt misunderstands how such variables evolve and, in doing so, severely curtails the scope for explaining the observed movements in consumption expenditure.

An empirical agenda is suggested that has the capacity to highlight the relevance and importance of neo-Schumpeterian economics by applying it in contexts of strong contemporary interest, such as economic growth. By interpreting existing evidence and respecifying models from such a perspective, it might become possible to alter the theoretical frameworks that many economists routinely apply. The consignment of economic history into the anonymity of the 'DGP' can be reversed and the forces discussed in neoclassical economics can be returned to their proper Marshallian context, namely, as a body of price theory that helps us understand the forces that move historical tendencies around in the short period and provide fuel for economic aspirations, in the Austrian sense, that impel us to plan and act in the longer term.

NOTE

* I would like to thank all participants of the EMAEE conference in Vienna, 13–15 September 2001, who commented on a previous draft of this chapter. However, the usual caveat applies.

REFERENCES

Aghion, P. and Howitt, P. (1998), *Endogenous Growth Theory*, Cambridge, MA: MIT Press.

Allen, P.M. (2001), 'Knowledge, ignorance and the evolution of complex systems', in J. Foster and J.S. Metcalfe (eds), *Frontiers of Evolutionary Economics: Competition, Self-Organization and Innovation Policy*, Cheltenham, UK and Northampton, USA: Edward Elgar.

Baba, Y., Hendry, D.F. and Starr, R.M. (1992), 'The demand for M1 in the U.S.A., 1960–1988', *Review of Economic Studies*, **59**, 25–61.

Bewley, R. and Griffiths, W.E. (1999), 'A forecasting comparison of classical and Bayesian forecasting methods for modelling logistic diffusion', *UNE Working Papers in Econometrics and Applied Statistics*, No. 107, September.

Dixon, R. (1994), 'The logistic family of discrete dynamic models', in J. Creedy and V. Martin (eds), *Chaos and Non-Linear Models in Economics*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Doornik, J.A. and Hendry, D.F. (1994), *PC GIVE 8.0 An Interactive Econometric Modelling System*, London: International Thomson Publishing.

Dopfer, K. (1988), 'Classical mechanics with a ethical dimension: Professor Tinbergen's economics', *Journal of Economic Issues*, **22**, 675–706.

Durlauf, S.N. (2001), 'A manifesto for a growth econometrics', *Journal of Econometrics*, **100**, 65–9.

Fagerberg, J. (2000), 'Technical progress, structural change and productivity growth: a comparative study', *Structural Change and Economic Dynamics*, **11**, 393–411.

Foster, J. (1987), *Evolutionary Macroeconomics*, London: Unwin Hyman.

Foster, J. (1991), 'Econometric methodology in an environment of evolutionary change', in S. Metcalfe and P. Saviotti (eds), *Evolutionary Theories of Economic and Technological Change*, Chur: Harwood Press.

Foster, J. (1992), 'The determination of Sterling M3, 1963–88: an evolutionary macroeconomic approach', *The Economic Journal*, **102**, 481–96.

Foster, J. and Wild, P. (1996), 'Economic evolution and the science of synergetics', *Journal of Evolutionary Economics*, **6**, 239–60.

Foster, J. and Wild, P. (1999a), 'Econometric modelling in the presence of evolutionary change', *Cambridge Journal of Economics*, **23**, 749–70.

Foster, J. and Wild, P. (1999b), 'Detecting self-organisational change in economic processes exhibiting logistic growth', *Journal of Evolutionary Economics,* **9**, 109–33. Reprinted in U. Cantner, H. Hanusch and S. Klepper (eds), *Economic Evolution and Complexity*, New York: Physica-Verlag, 2000.

Hannan, M. and Freeman, J. (1989), *Organisational Ecology*, Cambridge, MA: Harvard University Press.

Hendry, D.F. (1980), 'Econometrics – alchemy or science?' *Economica*, **47**, 387–406.

Hendry, D.F. (1995), *Dynamic Econometrics*, Oxford: Oxford University Press.
Hendry, D.F. (2000), 'On detectable and non detectable structural change', *Structural Change and Economic Dynamics*, **11**, 45–65.
Jones, C.I. (1998), *Introduction to Economic Growth*, New York: Norton.
Lawson, T. (1997), *Economics and Reality*, London: Routledge.
Leamer, E.E. (1983), 'Let's take the con out of econometrics', *American Economic Review*, **73**, 31–43.
Morgan, M. (1989), *The History of Econometric Ideas*, Cambridge: Cambridge University Press.
Nelson, R. (1997), 'How new is new growth theory?', *Challenge*, **40**, 29–58.
Nelson R. and Winter, S. (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Belknap Press.
Pagan, A.R. (1987), 'Three econometric methodologies: a critical appraisal', *Journal of Economic Surveys*, **1**, 3–24.
Pagan, A.R. (1994), 'Three econometric methodologies: an update', in L.T. Oxley, C.J. Roberts, D.W.R. George and S.T. Sayers (eds), *Surveys in Econometrics*, Oxford: Blackwell.
Potts, J. (2000), *The New Evolutionary Microeconomics: Complexity, Competence and Adaptive Behaviour*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Sarkar, J. (1998), 'Technological diffusion: alternative theories and historical evidence', *Journal of Economic Surveys*, **12**, 131–76.
Schumpeter, J.A. (1933), 'The common sense of econometrics', *Econometrica*, **1**, 5–12.
Scott, M.F. (1989), *A New Theory of Economic Growth*, Oxford: Clarendon Press.
Sims, C.A. (1980), 'Macroeconomics and reality', *Econometrica*, **48**, 1–48.

3. Random walks and non-linear paths in macroeconomic time series: some evidence and implications

Franco Bevilacqua and Adriaan van Zon

1. INTRODUCTION

The aim of this chapter is to identify the nature of the dynamics of macroeconomic time series. When time series are characterized by zero autocorrelation for all possible leads and lags, the issue of distinguishing between deterministic and stochastic components becomes an impossible task when linear methods are used (Hommes 1998).

This impasse arises because linear methods are appropriate to detect regularities in time series like autocorrelations and dominant frequencies (Conover 1971, Oppenheim and Schafer 1989), while fluctuations in real economic time series are generally characterized by zero autocorrelation and no dominant frequency. Economic fluctuations seem really similar to background noise, which does not possess dominant frequencies and each noise impulse is not serially correlated. The spectral analysis of economic fluctuations, seemingly as complex as noise, has led many economists to consider fluctuations as identically independently distributed (i.i.d.) events.

As a matter of fact the i.i.d. hypothesis is an obvious necessity for all linear models to describe, at least approximately, the irregularities in the observed data. In the past two kinds of linear economic models based on the i.i.d. hypothesis in the residuals have been presented. In the first model, known as the *deterministic trend* model, variables evolve as a function in time along a linear trend. In the second model (the *stochastic trend* model), variables evolve as a function of their forgoing values and a shock shifts the value of the variable from the lagged value (Rappoport and Reichlin 1989). In this second case any shock evidently affects the value of the variable at all leads and, therefore, it has a persistent effect. Moreover the time series is entirely determined by the occurrence of all past shocks (Fuller 1999, Maddala and Kim 1998).

Following the seminal article by Nelson and Plosser (1982), the empirical evidence in the last 20 years has contradicted the linear trend models.

The stochastic trend model put forward by Nelson and Plosser seemed, instead, not to be contradicted by empirical results.

In this chapter the Nelson and Plosser model will be called in question because it is based on the hypothesis that fluctuations are i.i.d. while they are not. The i.i.d. hypothesis, in our opinion, obscures existent non-linearities that may be endogenized in non-linear models.

This chapter is organized as follows. In Section 2 the main stylized facts offered by the recent linear econometric analysis are presented. In Section 3 it is shown how neoclassical economic theory can be fully consistent with recent econometric results. In Section 4 we put forward the hypothesis that non-linearities of the system may be a deterministic cause of the irregularities in economic time series and we introduce a procedure, based on recent non-linear signal processing techniques, that allows us to identify the existence of non-linearities in the system and, it is hoped, to filter out non-linearities (signals) from truly i.i.d. components (noise). In Section 5 we present results obtained using artificial non-linear and autoregressive models; in particular we use the arsenal of tools from non-linear dynamics to identify the hidden deterministic structure that underlies the time series. In Section 6 we present results obtained using non-linear metric techniques applied to monthly seasonally adjusted time series of some real macroeconomic time series of the USA (industrial production, employment, consumer price index, hourly wages, and so on). The common result that stands out from this analysis is that all the time series we have analyzed are also characterized by non-random structures in the residuals and therefore the i.i.d. hypothesis is simply inconsistent with facts. The choice of assuming the residual components as random neglects the existence of a complex phenomenon. Instead, it is even theoretically possible to reduce any stochastic component that perturbs the system unpredictably and thus highlight the non-linear deterministic component. In Section 7 we conclude by showing some theoretical implications that we can infer from our empirical results about real business cycle theory grounded on stochastic components with persistent effects.

2. EMPIRICAL EVIDENCE

In the last 20 years we have witnessed huge progress in the statistical and econometric analysis of time series which has given economists a far more profound knowledge about the relations between economic variables. The discovery and the realization that time series do not show any tendency to evolve along a deterministic log-linear growth trend and the cyclical reversible components, assumed in classical econometrics, do not exist at all has deeply marked the direction of the empirical research in the last two decades.

Recent econometric works have provided a solid empirical basis that is in contrast to the theoretical results of the early neoclassical growth models *à la* Solow (1956) and the business cycle models *à la* Lucas (1972, 1977 and 1980) based on monetary disturbances with transitory effects. Nelson and Plosser (1982) have provided empirical evidence for the theoretical alternative to the real business cycle, despite the conventional wisdom of classical econometrics that assumed *ex ante* stationarity for all the economic variables. Nelson and Plosser have shown that many macroeconomic time series[1] are not stationary at all, and the stationary stochastic models developed in the 1970s do not actually have any empirical foundation.[2]

On the contrary, Nelson and Plosser have shown that the irregularity present in macroeconomic time series could simply be explained by the introduction of random shocks with persistent effects as happens in unit root processes.[3]

These results were in sharp contrast with the classic econometric works, which affirmed that the irregularity in economic time series was due to transitory shocks, and has been crucial in moving the direction of research towards the theory of the real business cycle.

The acknowledged contribution of the Nelson and Plosser work was the discovery of the non-stationarity in the time series and the absence of any deterministic trend. More importantly, the introduction of random external shocks as the unique generator of the irregularity in the behavior of economic systems did not contradict the results put forward by a modern version of neoclassical theory: real business cycle theory. Indeed, without the injection of external shocks, time series would move exactly in the direction that the neoclassical theory predicts. However, in the presence of external shocks, economic systems move irregularly in the way that is described by real business cycle models (Prescott 1998).

In this chapter we try to move a step forward, starting from this empirical evidence. Our aim is to identify the process that generates the non-stationarity in time series without stating *ex ante*, contrary to Nelson and Plosser, that the non-stationarity is the direct consequence of a stochastic process. Actually there may be many possible non-linear deterministic alternatives to the stochastic explanation of the non-stationarity in time series.

Treating economic fluctuations as an endogenous non-linear process, and therefore an object of analysis, may contribute to a better understanding of the temporal evolution of time series. Our purpose is to understand the dynamics of fluctuations as the evolution of the system may depend entirely on them. We believe that assuming fluctuations as i.i.d. variables equivalent to noise is basically wrong since, as we shall see in Section 6, residuals are characterized by a structure that is very different from noise and even from any other kind of random variable. These results will lead us to conclude that it is feasible to discover deterministic laws that shape the underlying non-linear structures.

2.1 Recent Results from the Unit Root Literature

Many recent related works have been published after the Nelson and Plosser paper and their results differ mainly in respect of the test function that has been used in the verification of the non-stationarity hypothesis.

Some papers simply confirm that the non-stationarity of economic time series is a recurrent characteristic in many countries. Similarly to Nelson and Plosser, Lee and Siklos (1991) found that macroeconomic time series for Canada are not stationary. Mills (1992) obtained basically the same results for the UK, McDougall (1995) for New Zealand, Rahman and Mustafa (1997) for the Asian countries, Sosa for Argentina (1997), Gallegati (1996) and de Haan and Zelhorst (1994) for Italy.

The macroeconomic variables that are more frequently analyzed are GDP, GNP, GDP per capita and GNP per capita, industrial production, employment, unemployment rate and the consumer price index. Occasionally other variables such as savings (Coakley et al. 1995), investments (Coorey 1991, Coakley et al. 1995), wages (Coorey 1991), exchange rates (Durlauf 1993, Parikh 1994, Wu and Crato 1995, Serletis and Zimonopoulos 1997, Weliwita 1998), money and velocity of money (Al Bazai 1998, Serletis 1994) have been analyzed.

All these studies pointed out that almost every time series in any country is characterized by the presence of a unit root, or equivalently by a stochastic process like a random walk.[4] The one exception to the existence of a unit root in macroeconomic time series is the unemployment rate. This non-conformity was first noticed by Nelson and Plosser and has been confirmed by the majority of unit roots researchers afterwards.[5]

In Table 3.1 we list the main works that have ascertained the existence of a unit root in macroeconomic time series. For each author we mark with a plus sign (+) the variable that was found to follow a random walk, and with an equals sign (=) the variable for which the results were mixed.

2.2 The Broken Trend Hypothesis

Rappoport and Reichlin (1986, 1988, 1989) put forward the hypothesis that there could exist a broken deterministic trend that cannot be identified by the Dickey–Fuller test. Rappoport and Reichlin showed that in the case of

Table 3.1 Main works ascertaining existence of a unit root in macroeconomic time series

Authors	Y	Y/N	Y_i	Y_a	GNP	GNP/N	Exp	S	I	E	U	r	e	p	m	v	Country
Nelson and Plosser (1982)																	USA
Lee and Siklos (1991)																	Canada
McDougall (1995)																	New Zealand
Franses and Kleibergen (1996)					+	+				+	=			+			USA
Wells (1997)																	USA
Nunes et al. (1997)																	USA
Coorey (1991)									+								USA
Mills (1992)																	UK
Durlauf (1993)	+																USA
de Haan and Zelhorst (1994)																	Italy
Sosa (1997)																	Argentina
Bannerjee et al. (1992)	+				+	+				+	+			+			7 OECD, Japan
Fung and Lo (1992)													+				USA
Parikh (1994)																	Japan, UK, Germany
Mocan (1994)					+												USA
Gamber and Sorensen (1994)											+						USA
Haslag et al. (1994)																	USA
Serletis (1994)																+	USA
Bresson and Celimene (1995)																	Caribbean
Dolado and Lopez (1996)											+						Spain
Leybourne and McCabe (1999)																	USA

Wu and Crato (1995)											New Zealand
Serletis and Zimonopoulos (1997)								+			17 OECD
Weliwita (1998)											Asia
Gallegati (1996)	+		+	+							Italy
Rahman and Mustafa (1997)	+				+						Asia
Bohl (1998)		+									G7
Al Bazai (1998)										+	Saudi Arabia
Choi and Yu (1997)									+		OECD
Coakley et al. (1995)						+	+				OECD
Osborn et al. (1999)			+								USA

Notes:
Y = GDP, Y/N = GDP per capita, Y_i = industrial production, Y_a = agriculture, Exp = exports, S = savings, I = investments, E = employment, U = unemployment rate, r = interest rate, e = exchange rate, p = consumer price index, m = money, v = velocity of money

a broken deterministic trend the Dickey–Fuller test produces spurious results, since it is incapable of rejecting a false null hypothesis (the unit root hypothesis). Rappoport and Reichlin have moreover revealed empirical evidence concerning the existence of a broken trend in many macroeconomic time series. They indeed rejected the hypothesis of a random walk for many real variables (such as industrial production, real GNP, real per capita GNP and money supply), though not for all of them.[6]

Perron (1989) as well as Rappoport and Reichlin showed that, when fluctuations are stationary along a broken trend, the Dickey–Fuller test is not able to reject the unit root hypothesis. Perron developed a test that allows to reject the unit root null hypothesis if the series is characterized by a broken trend. He applied his test to the same time series of the USA that was used by Nelson and Plosser, after he had arbitrarily assigned the date at which the structural break occurred. Perron concluded that the null unit root hypothesis could be rejected also at a high confidence level for almost all the time series.

Similar results were obtained by Raj (1992) for the macroeconomic time series of Canada, France and Denmark, by Rudebusch (1992) for England, by Linden (1992) for Finland, by Wu and Chen (1995) for Taiwan and by Soejima (1995) for Japan.

Other authors also looked for a broken trend in specific time series. Diebold and Rudebush (1989), Duck (1992), Zelhorst and de Haan (1993), Ben-David and Papell (1994), Alba and Papell (1995) and McCoskey and Selden (1998) have found a broken trend for the GDP in many countries. Alba and Papell (1995) for GDP per capita and Li (1995), Gil and Robinson (1997) found similar results for industrial production, Simkins (1994) for wages in eight OECD countries and McCoskey and Selden (1998) for the G7 countries, Raj and Slottje (1994) for the US income distribution, Culver and Papell (1995), Leslie et al. (1995), and MacDonald (1996) for exchange rates. Given these results, we could check whether the broken trend hypothesis also explains the dynamics of unemployment rate better than the unit root hypothesis. However, Nelson and Plosser had already found that the US unemployment rate tended to be stationary, and the works by Hansen (1991), Li (1995), Leslie et al. (1995), Song and Wu (1997, 1998), Gil and Robinson (1997) and Hylleberg and Engle (1996) simply confirm the empirical evidence presented by Nelson and Plosser.

In Table 3.2 we present the main works that support the hypothesis of a broken trend in macroeconomic time series. For each author we mark with a minus (−) sign the variable that was found stationary along a broken trend.

Criticisms of both the broken trend and the unit root hypothesis have been put forward by several authors. Zivot and Andrews (1990, 1992)

Table 3.2 Main works supporting hypothesis of a broken trend in macroeconomic time series

Authors	Y	Y/N	Y_i	YD	Gnp	Gnp/N	E	U	w	r	e	p	Country
Perron (1989)													USA
Rudebusch (1990)													Australia
Raj (1992)													USA, Canada, France, Denmark
Rudebusch (1992)													Germany
Linden (1992)					–	–	–	–				–	Finland
Wu and Chen (1995)													Taiwan
Soejima (1995)													Japan
Lee (1996)													USA
Lumsdaine and Papell (1997)													USA
Diebold and Rudebusch (1989)													USA
Duck (1992)	–												9 countries
Ben-David and Papell (1994)													16 countries
Gokey (1990)													8 OECD
Capitelli and Scjlegel (1991)													6 countries
Leslie et al. (1995)										–			UK
Wu and Zhang (1996)													OECD
Moosa and Bhatti (1996)													Asia
Hansen (1991)													UK
Song and Wu (1997)								–					48 US states
Song and Wu (1998)													OECD
Hylleberg and Engle (1996)													OECD

Table 3.2 (Continued)

Authors	Y	Y/N	Y_i	YD	GNP	GNP/N	E	U	w	r	e	p	Country
Simkins (1994)							–		–				USA
Raj and Slottje (1994)				–									USA
Caselli and Marinelli (1994)					–								Italy
Culver and Papell (1995)											–		OECD
Wu (1996)													USA
Li (1995)			–										USA
Gil and Robinson (1997)													USA
Alba and Papell (1995)	–	–											Newly industrializing countries
Fleissing and Strauss (1997)									–				G7
McCoskey and Selden (1998)	–												OECD
Cheung and Chinn (1996)					–								USA
Dolmas et al. (1999)		–											USA

Notes:
Y = GDP, Y/N = GDP per capita, Y_i = industrial production, YD = income distribution, E = employment, U = unemployment rate, w = wages, r = interest rate, e = exchange rate, p = consumer price index

estimate the position in time of the structural break and find that the existence of the broken trend is not that clear in many of the time series that were analyzed by Perron. Cushing and McGarvey (1996) found that the fluctuations in the macroeconomic time series are more persistent compared to what stationary models indicate, but they are also less persistent than unit root models suggest. Mixed results were also obtained by Leybourne et al. (1996) for many US macroeconomic time series, Krol (1992) for the production of many US sectors, and Crosby (1998) for Australian GDP.

It seems therefore that not every time series is characterized by a unit root. What does this suggest? Are time series generated by a deterministic process or by chance? This issue has not been well formulated either in the unit root or in the broken trend literature. The problem is that the idea according to which a non-stationary process is a random walk process was implied in most of these studies. As we will see in Section 4, not all the non-stationary processes follow a random walk. Indeed, there may exist many deterministic non-linear processes that are not stationary and become stationary after differentiating with respect to time.

Since the results obtained by the broken trend literature are still open to discussion in the sense that the studies hitherto published do not lead to a general rejection of the random walk hypothesis, we question whether the broken trend hypothesis provides the ultimate explanation of the nature of economic time series. Moreover, as will be shown in the next section, the random walk hypothesis has the great advantage that it may be theoretically fully consistent with the neoclassical framework once it is assumed that real changes occur randomly.

3. THE LINK BETWEEN NEOCLASSICAL GROWTH THEORY AND THE UNIT ROOT LITERATURE

King et al. (1988b) showed that growth theory, which assumes steady growth, may be consistent with the highly irregular behavior of economic time series.

They considered a one-commodity Solow (1956) and Swan (1956) model. The production function, the capital accumulation equation and the resource constraint are:

$$\begin{gathered} Y_t = A_t K_t^{1-\alpha} (NX_t)^{\alpha} \qquad 0 < \alpha < 1 \\ K_{t+1} = I + (1-\delta)K_t = sA_tK_t^{1-\alpha} (NX_t)^{\alpha} + (1-\delta)K_t \\ L_t + N = 1 \\ C_t + I_t = Y_t, \end{gathered}$$

where Y_t is the output at time t, K_t is the capital stock available at time t, s the saving rate, N is the labor input that is assumed constant at all time t, A_t is a multiplier factor and its change corresponds to temporary changes of total factor productivity, X_tN is the *effective* labor units and changes of X_t modify permanently the performance of the system, and C_t is the consumption at time t.[7]

Assume constant returns to scale in the production function, and *constant* labor-augmenting technical change rate $\Delta(x)/(x)$. The dynamic equation for the capital stock may be rewritten as:

$$\Delta K_t = sA_tK_t^{1-\alpha}(NX_t)^{\alpha} - \delta K_t \rightarrow \frac{\Delta K_t}{N} = \frac{sA_tK_t^{1-\alpha}(NX_t)^{\alpha} - \delta K_t}{N}$$

$$\Delta k_t = sA_tk_t^{1-\alpha}N^{1-\alpha}N^{\alpha-1}X_t^{\alpha} - \delta k_t, \text{ where } k_t = \frac{K_t}{N}.$$

$$\frac{\Delta k_t}{k_t} = \frac{sA_tk_t^{1-\alpha}(X_t)^{\alpha} - \delta k_t}{k_t} = \frac{sA_tk_t^{1-\alpha}(X_t)^{\alpha}}{k_t} - \delta = \gamma,$$

where γ is the growth rate of the capital per capita.

If

$$\frac{sA_tk_t^{1-\alpha}(X_t)^{\alpha}}{k_t} > \delta, \frac{\Delta k_t}{k_t} > 0,$$

capital per capita grows.

Conversely if

$$\frac{sA_tk_t^{1-\alpha}(X_t)^{\alpha}}{k_t} < \delta, \frac{\Delta k_t}{k_t} < 0,$$

capital per capita decreases.

In steady state

$$\frac{\Delta k_t}{k_t} = 0 \text{ and } \frac{A_tk_t^{1-\alpha}(X_t)^{\alpha}}{k_t} = A_tk_t^{-\alpha}(X_t)^{\alpha} = \frac{\delta}{s} \text{ is constant.}$$

In order that $A_tk_t^{-\alpha}(X_t)^{\alpha}$ is constant over time, k_t and X_t must grow at the same rate γ. The output per capita is $y_t = A_tk_t^{1-\alpha}(X_t)^{\alpha} = kA_tk_t^{-\alpha}\ (X_t)^{\alpha}$; in steady state, being $A_tk_t^{-\alpha}(X_t)^{\alpha} = \delta/s$, also y_t grows at the same rate of k, γ. Consumption per capita is $c = (1 - s)y$ and grows at the same rate γ over time. In this sense, macroeconomic variables follow a (linear) deterministic trend.

This view was in sharp contrast with the empirical evidence from Nelson and Plosser (1982), who showed that the existence of a stochastic trend should not be neglected. However, it is very easy to make stochastic the basic version of the deterministic neoclassical model.

To do that, we consider that the labor-augmenting technical change occurs *stochastically* as a random walk.

We have:

$$X_\tau = X_0 \gamma^\tau e^{\sum_{t=0}^{\tau} \varepsilon_{t-i}} \rightarrow \ln X_\tau = \ln X_0 + \tau \ln\gamma + \sum_{t=0}^{\tau} \varepsilon_{t-i},$$

where $\sum_{t=0}^{\tau} \varepsilon_{t-i}$ represent permanent shifts of $\ln X_\tau$ which are not reabsorbed by the internal dynamics of the system.

Given the dynamic equation for capital accumulation, in steady state,

$$\frac{\Delta k_\tau}{k_\tau} = 0 \text{ and } \frac{A_\tau k_\tau^{1-\alpha}(X_\tau)^\alpha}{k_\tau} = A_\tau k_\tau^{-\alpha} (X_\tau)^\alpha = \frac{\delta}{s} \text{ is constant.}$$

In order that $A_\tau k_\tau^{-\alpha} (X_\tau)^\alpha$ is constant over time, k_τ and X_τ must grow at the same stochastically by

$$\gamma^\tau e^{\sum_{t=0}^{\tau} \varepsilon_{t-i}}:$$

$$\ln k_\tau = \ln k_0 + \tau \ln\gamma + \sum_{t=0}^{\tau} \varepsilon_{t-i}$$

The output per capita is $y_\tau = A_\tau k_\tau^{1-\alpha} (X_\tau)^\alpha = k A_\tau k_\tau^{-\alpha} (X_\tau)^\alpha$ in steady state, being $A_\tau k_\tau^{-\alpha} (X_\tau)^\alpha = \delta/s$, y_τ grows also by

$$\gamma^\tau e^{\sum_{t=0}^{\tau} \varepsilon_{t-i}}:$$

$$\ln y_\tau = \ln y_0 + \tau \ln\gamma + \sum_{t=0}^{\tau} \varepsilon_{t-i}$$

Consumption per capita is $c_\tau = (1-s) y_\tau$ and grows by

$$\gamma^\tau e^{\sum_{t=0}^{\tau} \varepsilon_{t-i}}:$$

$$\ln c_\tau = \ln c_0 + \tau \ln\gamma + \sum_{t=0}^{\tau} \varepsilon_{t-i}$$

In this sense, macroeconomic variables follow a *stochastic* trend where all the dynamics is driven by additive random innovations. Most of the empirical studies confirm that: (1) macroeconomic variables follow a stochastic trend, that is, a random walk; (2) macroeconomic variables co-evolve together; that is, they are co-integrated. This is exactly what occurs in the stochastic formulation of the neoclassical model. In fact, the above equations may be equivalently rewritten in terms of an AR(1) process:

$$\ln X_t = \ln X_{t-1} + \ln \gamma + \varepsilon_t$$
$$\ln k_t = \ln k_{t-1} + \ln\gamma + \varepsilon_t$$

$$\ln y_t = \ln y_{t-1} + \ln\gamma + \varepsilon_t$$
$$\ln c_t = \ln c_{t-1} + \ln\gamma + \varepsilon_t,$$

where all the economic variables depend on their previous value, on the average growth rate plus a non-transitory stochastic error term.

What is implicit in the stochastic version of the neoclassical model is that the economic system is essentially stable. In fact, if time series follow a random walk and we remove random innovations, we have a stationary stable system. In absence of technical change the system would never change, except for the occurrence of other exogenous shocks, for instance a change in preferences.

If the term $\Sigma_{t=0}^{T} \varepsilon_{t-i}$ were not random, what would be the consequences for economic theory? The first consequence would be that, understanding the deterministic non-linear dynamics, we could make a better prediction than simple AR-like models, since the best predictor for the residual in the AR models cannot be but its mean value. The second consequence would be that economic systems might be intrinsically unstable, that is, also without the injection of exogenous random inputs the system could not be motionless. Moreover, just because real economic time series prove to be complex, seemingly random but containing some deterministic structure, they could be better forecasted and better controlled.

In the next section we raise the hypothesis that residuals might appear random while they are indeed generated by a deterministic system. Later on, in Section 6, we will test whether or not the residual component of an autoregresssive model is truly random, and we will find, to our surprise, that the hypothesis that $\Sigma_{t=0}^{T} \varepsilon_{t-i}$ is not truly random is indeed found in our inference.

4. THE NON-LINEAR HYPOTHESIS[8]

Twenty years after the publication of the Nelson and Plosser article, we now have two literature streams that debate the nature of the time series: one that underlines the existence of a random walk and one that asserts the complete linear (though with a break) determinism in the economic time series. We will show in Section 6 that the empirical evidence about the nature of economic time series can be clearer than that provided by either the unit root literature or the broken trend literature.

The procedure that will be used in Sections 5 and 6 to detect non-linearities consists of the following steps:

1. Select time series with a minimal number of observations. Brock et al. (1991) have proved that at least 400 observations would be a good start-

ing point, if not a necessary condition, for obtaining trustworthy results from the BDS test. It is therefore necessary to rely on seasonally adjusted monthly data for a sufficiently long period.[9] The time series we used are those of the USA and data were provided by the Bureau of Labor and Statistics and the Federal Reserve.[10]

2. Take the natural logs of the original time series if the time series tend to diverge exponentially.
3. Differentiate the time series once with respect to time, eventually remove linear autocorrelation in the residuals and check for stationarity via the augmented Dickey–Fuller test.
4. Calculate the level of spatio-temporal entropy[11] to measure the degree of disorder of the system. If the time series of the residual was generated by a random process, the level of entropy should be close to the maximal value. However, non-linear processes may also present a high degree of disorder and reach values of entropy close to that of white noise.[12] On the other hand we should expect a low level of entropy for processes that are deterministic and autocorrelated.[13] However, we should not overestimate the importance of the measure of entropy; in fact it does not allow us to distinguish a random process from a complex deterministic one and even between periodic cycles and linear trend. Nevertheless the measure of entropy may help us to better understand the complexity of a time series.
5. Calculate the values of the maximal Liapunov exponents that characterize the time series, to measure how fast nearby trajectories diverge over time. If the maximal Liapunov exponents turn out to be negative, this means that trajectories tend to converge to a stable fixed point. If it were zero, we would have found a limit cycle. If it were positive, the time series is either characterized by chaos or a random walk. We anticipate that the residuals of the linear models that explain economic time series are generally characterized by a positive maximal Liapunov exponent and a high level of entropy, and this indicates how difficult it might be to forecast economic time series in the long run.
6. Generate *Ruelle plots* (recurrence plots) to uncover, from the qualitative point of view, hidden structures in the time series.
7. Perform the BDS test to detect quantitatively and in a reliable way the existence of non-linearity in data.
8. Check results randomly by shuffling the time series and verify whether the results obtained by the BDS test applied on a randomly shuffled time series are indeed different from the results obtained by the BDS test on the original time series.[14] This verification is extremely important since, if the two results turn out to be different, it means that the

time order of the original time series is significant and there exists causality in the data.

5. RESULTS FROM ARTIFICIAL TIME SERIES

Before applying the described procedure to real time series, we present some results obtained from artificial time series, whose deterministic data generating process is known. We present some cases of deterministic systems whose dynamics is very similar to a random walk and we check whether the non-linear dynamics tools allow us to gain more information about the nature and the evolution of the time series. We will see that the information gain ensued from the numerical tools of non-linear time series analysis may be relevant and may lead us to consider the issues of dynamics from a very different perspective.

5.1 Trends

We consider first the most simple case: growth along a linear trend. We first check the results obtained with the Dickey–Fuller test when a linear time series grows deterministically with time. Thereafter we apply non-linear metric tools to see which other information may be obtained. The application of non-linear techniques to a linear system may not seem to be necessary, but this step will allow us to compare the information that can be obtained using linear statistics and non-linear dynamics tools.

In the trend stationary case, residuals have no persistent effects and the time series is stationary along a linear trend. If we consider the variable x_t as a linear function of time t: $x_t = x_0 + \phi t + \varepsilon_t$ where x_0 is the initial value (in our case it is equal to zero), ϕ is a parameter and ε_t is an i.i.d. variable. Running the Dickey–Fuller test we should reject correctly the null hypothesis of a unit root and the Durbin–Watson statistics, DW, should be around 2 (when $DW \simeq 2$, residuals have no serial correlation).

Suppose that we are interested in studying the dynamics of a variable that could be the GDP, y_t. We assume that GDP grows at the yearly rate $g = 2$ per cent:

$$y_t = y_0(1+g)^t \rightarrow \ln y_t = \ln y_0(1+g)^t \rightarrow \ln y_t = \ln y_0 + t\ln(1+g)$$

Suppose that $\ln y_t$ is perturbed by a i.i.d. exogenous shock ε:[15]

$$\ln y_t = \ln y_0 + t\ln(1+g) + \varepsilon_t.$$

Table 3.3 Rejection of the null hypothesis of unit root

	Deterministic trend	Random walk	'Tent map' walk	'Rossler' walk
ADF statistic	−21.3**	−1.98	−1.79	−67.53**
D–W statistic	2.00	1.99	2.00	0.09**
Entropy of residuals	90%	90%	78%	15%
BDS statistic	−1.28	−1.55	99.2**	355.0**

Note: * and ** denote significance at the 5% and 1% levels.

Set $\ln y_t = x_t$ and $\ln(1+g) = \phi$ we obtain:

$$x_t = x_0 + \phi t + \varepsilon_t \text{ where } g = 0.02 \text{ and } \phi = 0.02.$$

Applying the Dickey–Fuller test we decidedly reject the null hypothesis of unit root (Table 3.3). The Dickey–Fuller test turned out to be −21.3 while the critical value at 5 percent significance level is −3.41. For values less than 3.41, the null hypothesis is rejected, as it is in this case. The Durbin–Watson statistics turned out to be close to 2, and this confirms that the residuals are not serially correlated. In this case, the Dickey–Fuller test was able to correctly reject the null hypothesis of a stochastic trend and to accept correctly the alternative hypothesis of a linear trend.

Let us now turn our attention to some qualitative and quantitative measurements obtained with non-linear dynamics tools. The value of *entropy* that characterizes the level of GDP is 0 percent, and this indicates that the time series is characterized by an almost null degree of disorder. In fact residuals are all concentrated around a linear trend, which represents a long-term equilibrium path. If we analyze the residuals, which were assumed to be i.i.d., the level of entropy turns out to be 90 percent, a value relatively close to the ideal limit of 100 percent of a purely random process (a value that is very difficult to reach in series generated by the simple algorithms of a random number generator). This indicates that the degree of disorder of a system characterized only by an i.i.d. variable is very high.

We have calculated the value of the maximal Liapunov exponent for the residuals, in order to measure the rate of sensitive dependence on initial conditions, that is, the rate of divergence of nearby initial states. It turned out to be positive (Table 3.5, row i.i.d. process) and so high that residuals follow a unpredictable dynamics. As we will see in Section 5.3, high values of the maximal Liapunov exponent and entropy are also typical of many non-linear systems.

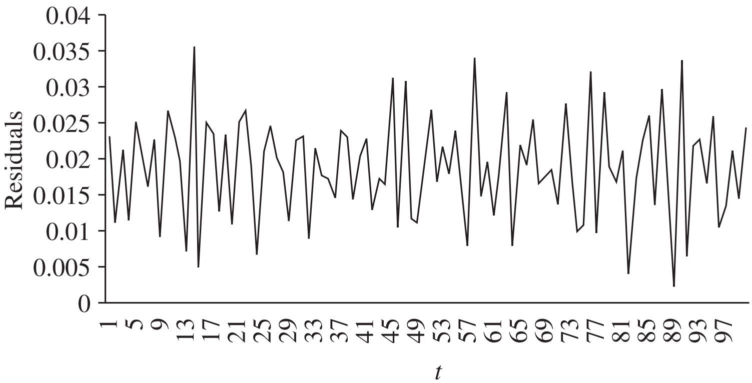


Figure 3.1 i.i.d. residuals (uniformly distributed)

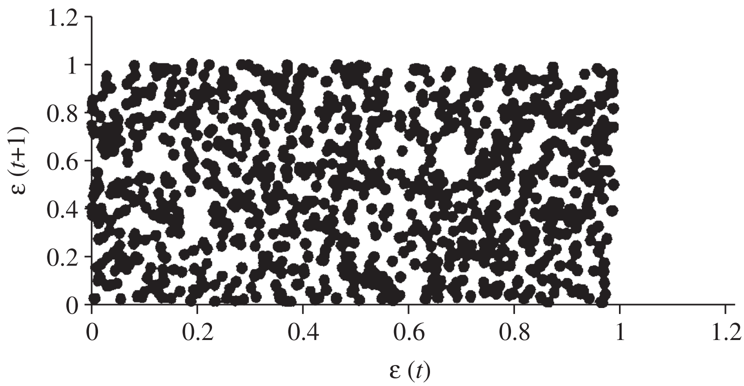


Figure 3.2 Phase plot of i.i.d. residuals

There are also qualitative visual devices that allow us to uncover complex structures in data and even to single out exceptional historical events. They are the *phase portraits* and the recurrence plots. The phase portrait is simply a graphical representation that plots the value $x(t)$ against $x(t-h)$. While in Figure 3.1 the residuals $\varepsilon(t)$ are plotted against time, they are plotted against $\varepsilon(t-1)$ in Figure 3.2.[16]

The recurrence plots by Eckmann et al. (1987) are a graphical tool for the qualitative analysis of time series based on phase portraits and allow us to uncover deterministic structures that could not be revealed by phase

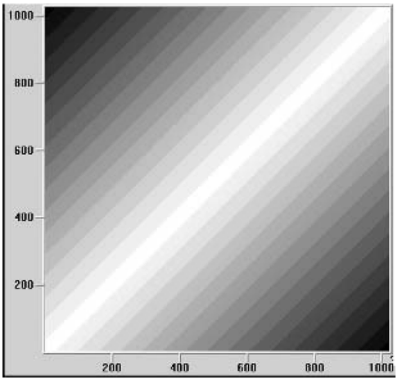


Figure 3.3 Recurrence plot of a linear trend

plots. In the simplest recurrence plots, the distances between observations are measured and marked by a grey tone. On the axis each point corresponds to a dated observation. The diagonal is the locus where $||x(t)-x(t-h)||=0$, where $h=0$ and the corresponding tone is white.

In the case of a deterministic trend the distance grows with the temporal distance of observations. The most distant observations are $x(0)$ and $x(T)$; hence the points $[x(0)-x(T)]$ and $[x(T)-x(0)]$ are marked by a black tone (Figure 3.3). The points along the parallels to the 45 degree line are characterized by the same grey tone and this indicates that the couples of observations that keep the same temporal distance are also characterized by the same spatial distance (represented by the same grey tone).

On the contrary, recurrence plots of i.i.d. residuals should neither present any continuous line between points nor particular areas characterized by the same grey tone. The fact that some nearly continuous lines may be noticed (Figure 3.4) is due to the random number generator, which is a mathematical algorithm and therefore does not produce purely unstructured time series. However Figure 3.4 shows much less structure than the Ruelle plot in Figure 3.3 and is close to the one of a purely i. i.d. process.

Actually, Ruelle plots may allow us to single out much more hidden structures when they compare embedded vectors[17] instead of single observations. Ruelle plots mark the distances between points[18] with a grey tone. If we choose $m=1$, we obtain Figures 3.3 and 3.4. If we had chosen different values of m, we would also have graphs similar to Figures 3.3 and 3.4. However, in other cases, especially in the case of chaotic systems, the choice of appropriate values for m allows us to uncover otherwise neglected structures.

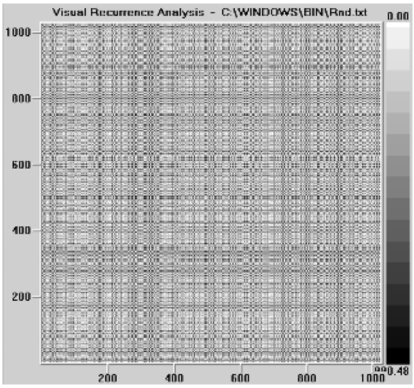


Figure 3.4 Recurrence plot of i.i.d. residuals

To discriminate a stochastic process from a process that contains a deterministic structure we apply the BDS test. The null hypothesis is that the time series is characterized by an i.i.d. process, while the alternative hypothesis is that the time series follows a non-linear law. Applying the BDS test to the residuals randomly generated by computer, we have found a value for the BDS function equal to -1.28 and a critical value of 1.96 at the 5 percent significance level. As expected, we accept the null i.i.d. hypothesis.

From this simple exercise we have obtained the following results:

- Using the Dickey–Fuller test, we have correctly concluded that the time series on levels is stationary and follows a deterministic trend.
- The entropy indicates that the time series of levels is stable and the time series of residuals is extremely unstable. The maximal Liapunov exponent of residuals is sharply positive, and this indicates that nearby trajectories diverge over time. Neither the values of entropy nor the maximal Liapunov exponent provide a definitive answer to the question regarding the nature of time series.
- Recurrence plots and phase portraits allow us to identify the existence of structures that are different from those of an i.i.d. process.
- The BDS test allows us to better appreciate the importance of the time order in time series, that is, to detect the existence of deterministic structures in time series. In this case we were not able to detect any deterministic structure in the residuals since there weren't any (except that of the random number generator algorithm).

5.2 Random Walks

We now analyze another limit case: the random walk. The random walk hypothesis is not generally rejected by the unit root literature and it is at the core of real business cycle theory.

In the random walk case, shocks, contrary to what happens in the case of deterministic trends, have persistent effects and accumulate over time, without being reabsorbed even partially in the future. The time series is not stationary, does not follow a linear trend, but can still grow in a quite similar way to the case of the deterministic trend. From a visual comparison between a series that grows like a random walk and a series that grows along a deterministic linear path, it is often not possible to distinguish the two time series. The Dickey–Fuller test serves to single out which of the two time series follows a random walk.

In a random walk process, the value of the variable x_t depends on its lagged value x_{t-1} and an i.i.d. shock ε_t:

$$x_t - x_{t-1} + \varepsilon_t$$

Suppose now that we are interested in the dynamics of a variable y that grows yearly at the average rate of 2 percent, as an effect of the accumulation of shocks:

$$\ln y_t = \ln y_{t-1} + \varepsilon_t \rightarrow \ln y_t - \ln y_{t-1} = \varepsilon_t \rightarrow \ln \frac{\gamma_t}{\gamma_{t-1}} = \varepsilon_t \rightarrow y_t = e^{\epsilon t} y_{t-1}$$

Plotting the log series against time, we would see a dynamics similar to the case of the deterministic trend (not shown here). It is not possible to determine which of the two time series is the random walk by a direct visual inspection alone. A growth trend exists, but it is a stochastic one.

To distinguish between a stochastic trend and a deterministic trend we apply the Dickey–Fuller test and, as we expected, we are not able to reject the unit root hypothesis. The value of the test function turned out to be -1.98, while the critical value is -3.41 at the 5 percent significance level (Table 3.3). Residuals turned out not to be serially correlated (the Durbin–Watson statistic is 1.99).

The entropy level, the maximal Liapunov exponent, the BDS test and Ruelle plots of the residuals are exactly the same as those obtained for the deterministic trend case. Inasmuch as the aim of non-linear dynamics is to detect complex structures in residuals, both in the case of stochastic growth and deterministic growth, residuals are stochastic and the tools of non-linear dynamics cannot be used to detect linear determinism. The suitable instrument to detect linear determinism is indeed the Dickey–Fuller test.

5.3 Non-linear Walks

5.3.1 Autoregressive tent map growth

We now apply the Dickey–Fuller test to an artificial time series where the value of the variable depends on its lagged value and a deterministic non-linear shock. We will apply the BDS test and other tools of non-linear dynamics to identify the deterministic structures that the Dickey–Fuller test is not able to detect.

Suppose that a time series is generated by the following deterministic law:

$$x_t = x_{t-1} + 0.04x_{t-1}\varepsilon_t \text{ with } \begin{array}{l} \varepsilon_t = 2\varepsilon_{t-1} \text{ for } \varepsilon_{t-1} < 0.5 \\ \varepsilon_t = 2(1-\varepsilon_{t-1}) \text{ for } \varepsilon_{t-1} < 0.5 \end{array}$$

This system is known as the *tent map* and it appeared in an *Economic Journal* article by Scheinkman (1990) and in a working paper of the University of Texas by Vastano and Wolf (1986). This peculiar system generates a chaotic time series which has the same statistical properties as a uniform distribution. Similarly to the random walk, $0.04\varepsilon_t$ has an average value equal to 0.02.[19]

A visual inspection of the generated time series x_t may be puzzling because x_t is very similar to a time series with either a deterministic or stochastic trend. In order to find out whether this system follows a stochastic or a deterministic trend we apply the Dickey–Fuller test and the unit root hypothesis cannot be rejected. In fact the value of the test function turned out to be −1.8 (Table 3.3), while the null hypothesis is rejected for values less than −3.4 at the 5 percent confidence level. The time series appears to be similar to the stochastic trend or to the deterministic linear trend. But we know that it is neither. The Durbin–Watson statistic turned out to be exactly equal to 2.00, and this indicates that residuals are not serially correlated. At this stage we would again apply the Dickey-Fuller test to the residuals to see whether they are stationary, and we would conclude that the process is autoregressive of order one with i.i.d. residuals.

This conclusion is only partly valid. The process is autoregressive of order one and therefore there is a unit root, but the residuals (shown in Figures 3.5 and 3.6) are deterministic and, knowing the law that generates the residuals, the process is perfectly predictable. In this case we must be very careful to read the results obtained with the Dickey–Fuller test. It suggests that it is not possible to reject the null hypothesis of the existence of a unit root, that is, the hypothesis of autoregressive process of order one. However, the residuals, as this case shows, can be non-stochastic. Consequently the Dickey–Fuller test is a tool that is not suitable for unveiling whether the series follows a deterministic law, except for the special case that the series follows a deterministic

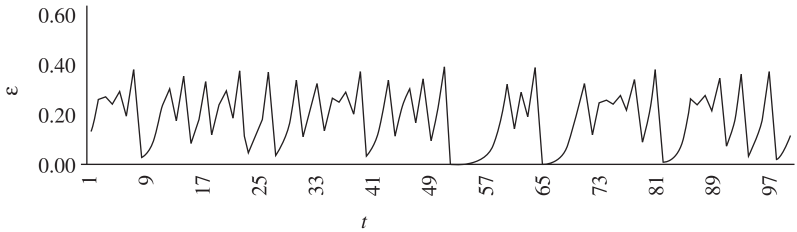


Figure 3.5 Tent growth residuals

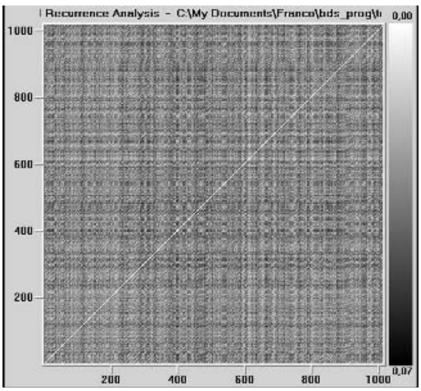


Figure 3.6 Recurrence plot of the 'tent map'

linear trend. The acceptance of a unit root hypothesis and the presence of non-serially correlated residuals does not authorize us to take the stochastic origin of the time series for granted.

From the values of entropy (78 percent) and the positive maximal Liapunov exponent we may infer that the system is nearly unpredictable. However, these characteristics are typical of both stochastic and chaotic processes. In order to infer the existence of non-linear structures we have performed the BDS test. The value of the BDS statistic (which asymptotically converges to normality)[20] turned out to be 99.2, and this allows us to reject the null i.i.d. hypothesis with a minimal probability of being mistaken.

5.3.2 Autoregressive Rossler growth

Consider the following system:

$$x_t = x_{t-1} + 0.02x_{t-1}(\tfrac{\varepsilon_t}{10} + 1),$$

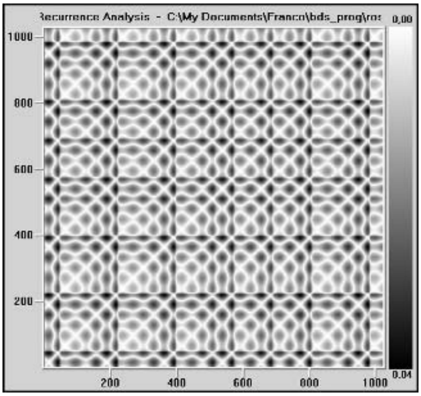


Figure 3.7 Recurrence plot of the 'Rossler' map

where $0.02(\frac{\varepsilon_t}{10} + 1)$ has an average value equal to 0.02, and ε_t is the result of a deterministic non-linear system that generates aperiodic (chaotic) cycles[21] (see Figure 3.8).

Applying the Dickey–Fuller test, we would reject the null hypothesis of autoregressive process of order one and accept the alternative hypothesis of a deterministic trend. The Dickey–Fuller statistic turned out to be −57.52, a value enormously greater than the respective critical value (−3.97 is the corresponding 5 percent critical value) (Table 3.3). The Durbin–Watson statistic turned out to be 0.09, and residuals are indeed serially correlated. Given these results, we would think that the time series follows a deterministic trend and fluctuations are cyclical with reversible effects. However, our model is autoregressive of order one, it does not follow a deterministic trend and the time series is entirely generated by fluctuations ε_t that have persistent effects.

The value of entropy of the residuals is 15 percent, and this low value implies that the system tends to preserve a certain stability over time. The maximal Liapunov exponent is positive and therefore the evolution of the system is sensitive with respect to its initial conditions, but since its value is close to zero, this suggests that the system is also cyclical. In fact it has aperiodic cycles; thus the system is also chaotic. The recurrence plots of the residuals (Figure 3.7), just like a simple graph against time (Figure 3.8), show a cyclical and a periodical dynamical structure.

The support for the existence of non-linear structures in the time series follows from the high value of the BDS statistic (Table 3.3). The null i.i.d. hypothesis is rejected. Although the BDS test was also able to detect correctly the existence of non-linear structures in the data in this case, we may

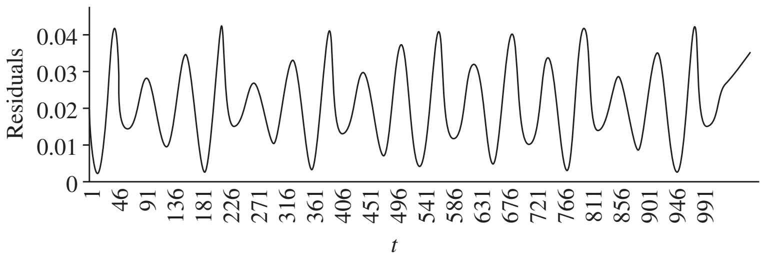


Figure 3.8 Rossler residuals

better appreciate its effectiveness when residuals are not serially correlated, as in the cases of the tent map and seasonally adjusted real time series.

6. EMPIRICAL EVIDENCE: THE US TIME SERIES

In the past 15 years the detection of non-linearities in real economic time series has turned out to be a very difficult task. The main problem is to apply the non-linear dynamic tools to time series that contain a sufficient number of observations. In order to reliably calculate the BDS test a quite high number of observations is needed. Around 400 observations are necessary to detect low-dimensional non-linearities. If we wish to discover more complex structures, we need an even higher number of observations. This is because the BDS test has a very low power for small finite samples. The application of the BDS test, as well as all the tools of non-linear dynamics based, as the BDS test, on the correlation dimension, to small samples may produce spurious results. In Section 5.2 the problem of spurious results did not arise since the sample was sufficiently large and consequently the power of the test was also high.

When we have a time series with a very limited number of observations it is necessary to use linear metrics; the use of non-linear dynamics tools would only produce wrong results. For instance, the frequency of observations for GDP is only quarterly and data are available starting from 1959. Although the Bureau of Economic Analysis is going to release these data from 1929, we could only have a maximum of 280 observations and this limitation would not allow us to prove the existence of a non-linear dynamics.[22]

Chavas and Holt (1991) have chosen to analyze a very specific time series which was already known to have a cyclical nature: the *Pork Cycle*. Chavas and Holt have shown the existence of aperiodic cycles in the quarterly time series of the US quantities and prices of pork meat from 1910 until 1984. Chavas and Holt have the great merit to have shown that fluctuations in time series may have a non-linear origin.

In the analysis that follows, we focus on some main real macroeconomic time series. We check whether it is possible to extract signals from the residuals that economic literature has assumed to be stochastic. What we want to ascertain is whether the residuals also contain a non-linear component together with a truly stochastic component. We try to find out whether important temporal linkages are present between residuals. We will attempt to falsify the results of rejection of the null i.i.d. hypothesis. We will proceed to a random shuffle of the time series in order to break any temporal link among data. Afterwards we will apply non-linear dynamics tools to the shuffled time series. If the results of non-linear tests on both the original and the shuffled time series are similar, it means that time linkages are not important and the time series is generated by a stochastic process; otherwise there is evidence that time cannot be ruled out and there exists a non-linear component.

6.1 Industrial Production

The time series for industrial production is certainly one of the most complete available. Data go back to 1919 and the frequency of observation is monthly.

Applying the Dickey–Fuller test[23] to the log of the observed values, we cannot reject the null hypothesis of a unit root (Table 3.4).

Then we estimated the following linear model that best fits the data:

$$Y(t) = 0.02 + 0.99\,Y(t-1) + 0.51[Y(t-1) - Y(t-2)] + 0{,}000029t + \hat{e},$$

where $Y(t)$ are the observed values of the industrial production in terms of value.[24] The Durbin–Watson statistic is 1.95, well within the acceptance range 1.89–2.10. This indicates that the estimated residuals are not serially correlated.

From the original series Y we focused on the estimated residuals $\hat{e}$. The residuals also appear to be characterized by a very complicated dynamics if we look at the entropy level (80 percent).

The calculus of the maximal Liapunov exponent depends on the parameter of the embedding dimension m. There exists a maximal Liapunov exponent for each value of m. The maximal Liapunov exponents are all positive for different values of m and this indicates a high sensitivity of the time series with respect to its initial conditions (Table 3.5).

The existence of a structured dynamics also seems to be corroborated by

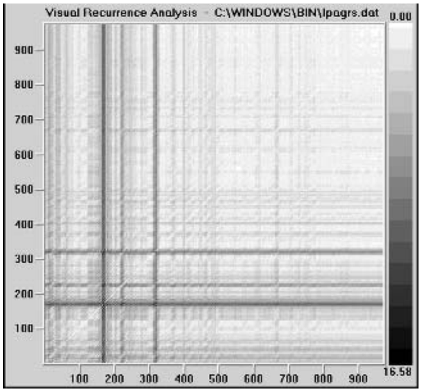


Figure 3.9 Recurrence plot, residuals, industrial production

the Ruelle plot[25] (Figure 3.9), where the presence of continuous lines is clear. In Figure 3.9 we can easily detect, without any *a priori* historical knowledge, the periods in which significant historical events have perturbed industrial production. From this recurrence plot we can realize that the first years of the 1920s, the years around 1933 and 1944, have been characterized by an anomalous dynamics. The embedded vectors represented by the single points around those dates show a big distance, marked with a dark shade, compared to nearly all the other vectors. Moreover, we can see that after the 400th embedded vector, the dynamics is more settled and also seems to repeat (see the bright area on the upper right). What is evident in Figure 3.9 is the existence of a structure that differs from a random walk (Figure 3.4).

To ascertain whether the time series is generated by a non-linear deterministic process we have applied the BDS test. The null i.i.d. hypothesis is strongly rejected (Table 3.4, column $W_{m,N}$). A similar test based on the same statistic of the BDS test is the dimension test (Table 3.4, column d_m). The correlation dimension d_m grows very slowly with m and tends to converge to a fixed value. This is typical of a process that is not guided by chance (Hommes 1998).[26]

If we randomize the order of the events of the original time series, we find that the values of the BDS test (column $W^{AF}_{m,N}$, Table 3.4) and the correlation dimension (column d^{AF}_m, Table 3.4) turn out to be very different from the values obtained using the original time series, and we correctly accept the null i.i.d. hypothesis for the shuffled time series. This is evidence that the time order of the residuals of the original time series is not random, and a temporal causality in the fluctuations exists.

We conclude that residuals in industrial production show a structure that cannot come from a mere linear stochastic process and therefore a non-linear explanation is necessary to understand the temporal causality of events. This result shows that there exists a clear non-linear structure in the estimated residuals, which in turn should be considered as truly signals and not as noise.

6.2 Empirical Analysis of Other Macroeconomic Time Series: Industrial Production in the Main US Sectors, Employment, Hourly Wages and the Consumer Price Index

A thorough analysis of each sector would be beyond the scope of this chapter, whose focus is on the existence of deterministic structures in macroeconomic time series. Shortly we summarize the results obtained analyzing some of the main US macroeconomic time series. We have restricted our analysis to the main sectors of the US economy,[27] employment, hourly wages and the consumer price index. Regarding the economic variables characterized by seasonal cycles we analyzed the seasonally adjusted time series. The frequency of observations is monthly. Data go back to 1947 for the transportation sector, industrial machinery and electrical machinery, 1967 for the hybrid high-tech sector (computers, semiconductors and communications), 1939 for employment, 1932 for hourly wages and 1913 for the consumer price index.

All the time series (log transformed), except employment, seem characterized by a unit root, since for most of them we are not able to reject the null i.i.d. hypothesis of the Dickey–Fuller test (Table 3.4) with high confidence levels (higher than 5 percent).[28] These results are qualitatively similar to those obtained by Nelson and Plosser. For all the time series, the estimated residuals of the linear model[29] that best fit the data turn out to be serially uncorrelated (the null hypothesis of the Durbin–Watson test is never rejected, even at a high confidence level for all the time series).

All the time series we analyzed are characterized by high entropy values (generally higher than 70 percent) that are typical of both chaotic and stochastic processes. For all the real time series we found positive values of the corresponding maximal Liapunov exponents (Table 3.5) and this result suggests that nearby trajectories diverge over time at a positive exponential rate. The interesting result is that all the real time series are characterized by a Liapunov exponent that is decidedly lower than that of an i.i.d. process, and lower than that of the tent map. This suggests that even if real time series have to be considered unpredictable in the long run, in the short run they are more predictable than an i.i.d. process and a deterministic process like the tent map.[30]

Table 3.4 Log-transformed time series

		Industrial production				Transportation equipment				Industrial machinery				Electrical machinery			
ADF statistic		−3.1				−3.9*				−3.8*				−2.8			
Estimated model		$Y_t = 0.02 + 0.99\,Y_{t-1} + 0.51\Delta Y_{t-1} + \varepsilon$				$Y_t = 0.13 + 0.96\,Y_{t-1} + \varepsilon$				$Y_t = 0.05 + 0.98\,Y_{t-1} + 0.09\Delta Y_{t-1} + 0.29\Delta Y_{t-2} + 0.26\Delta Y_{t-3} + \varepsilon$				$Y_t = 0.05 + 0.97\,Y_{t-1} + 0.17\Delta Y_{t-1} + \varepsilon$			
Durbin–Watson stat.		1.95				1.86				2.04				2.06			
Entropy of residuals		80%				73%				77%				Not available			
Parameters		BDS and dimension tests				BDS and dimension tests				BDS and dimension tests				BDS and dimension tests			
ε	m	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}
2σ	2	11.3**	0.13	0.9	0.15	10.4**	0.13	−0.3	0.17	4.1**	0.26	−0.1	0.27	4.1**	0.26	−0.1	0.27
2σ	4	15.1**	0.21	1.4	0.30	10.7**	0.23	0.1	0.33	3.8**	0.51	−0.5	0.55	3.8**	0.51	−0.5	0.55
2σ	6	16.2**	0.28	0.9	0.46	9.5**	0.32	0.1	0.49	4.1**	0.74	−0.7	0.83	4.1**	0.74	−0.7	0.83
σ	2	13.9**	0.21	0.6	0.25	11.8**	0.26	−0.7	0.34	3.5**	0.61	−0.7	0.49	3.5**	0.61	−0.7	0.49
σ	4	19.4**	0.32	0.7	0.49	13.0**	0.46	0.0	0.67	3.3**	1.2	−0.7	0.99	3.3**	1.2	−0.7	0.99
σ	6	22.5**	0.42	0.2	0.74	13.1**	0.65	0.0	1.00	3.8**	1.7	−0.8	1.50	3.8**	1.7	−0.8	1.50
0.5σ	2	15.5**	0.31	0.4	0.37	11.7**	0.45	−1.2	0.55	2.8**	0.9	−0.7	0.7	2.8**	0.9	−0.7	0.7
0.5σ	4	21.5**	0.50	0.4	0.74	13.9**	0.81	−0.4	1.09	2.9**	1.8	−0.7	1.4	2.9**	1.8	−0.7	1.4
0.5σ	6	27.3**	0.66	−0.2	1.12	15.3**	1.17	−0.0	1.63	3.6**	2.6	−0.5	2.1	3.6**	2.6	−0.5	2.1

Table 3.4 (Continued)

		High-tech				Employment				Hourly earnings				Consumer price index			
ADF statistic		0.6				−4.2**				−1.1				−0.8			
Estimated model		$Y_t = 1.00\,Y_{t-1} + 0.12\Delta Y_{t-1} + \varepsilon$				$Y_t = 0.16 + 0.97\,Y_{t-1} + 0.27\Delta Y_{t-1} + 0.27\Delta Y_{t-2} + \varepsilon$				$Y_t = 1.00\,Y_{t-1} + 0.20\Delta Y_{t-1} + 0.24\Delta Y_{t-2} + \varepsilon$				$Y_t = 1.00\,Y_{t-1} + 0.33\Delta Y_{t-1} + 0.16\Delta Y_{t-2} + 0.13\Delta Y_{t-3} + \varepsilon$			
Durbin–Watson stat		2.05				2.07				2.04				2.05			
Entropy of residuals		Not available				68%				71%				71%			
Parameters		BDS and dimension tests				BDS and dimension tests				BDS and dimension tests				BDS and dimension tests			
ε	m	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}	$W_{m,N}$	d_m	$W_{m,N}^{AF}$	d_m^{AF}
2σ	2	3.7**	0.13	0.0	0.17	10.8**	0.07	−1.0	0.09	12.2**	0.06	−1.6	0.08	8.2**	0.12	−1.4	0.13
2σ	4	4.0**	0.25	1.4	0.30	11.0**	0.13	−0.9	0.19	12.0**	0.11	−1.2	0.16	11.7**	0.21	−0.9	0.27
2σ	6	4.4**	0.36	2.2*	0.43	10.9**	0.18	−0.9	0.28	12.4**	0.15	−1.2	0.25	14.0**	0.27	−1.3	0.41
σ	2	4.7**	0.41	0.1	0.46	10.0**	0.17	−0.9	0.21	12.3**	0.17	−0.4	0.22	12.3**	0.26	−1.4	0.30
σ	4	6.3**	0.75	2.1*	0.87	12.1**	0.29	−0.1	0.41	15.4	0.28	−0.3	0.43	18.6**	0.43	−0.3	0.60
σ	6	7.9**	1.05	2.8**	1.25	14.6**	0.37	0.2	0.62	18.3**	0.37	−0.4	0.65	25.3**	0.56	0.2	0.90
0.5σ	2	4.9**	0.57	−0.1	0.63	9.1**	0.34	−0.3	0.40	12.9**	0.33	0.2	0.40	15.3**	0.43	−1.1	0.50
0.5σ	4	7.2**	1.05	1.9	1.20	14.7**	0.57	0.4	0.79	19.0**	0.56	−0.3	0.82	24.9**	0.72	−0.3	1.00
0.5σ	6	9.5**	1.47	2.4**	1.74	22.2**	0.75	0.2	1.19	26.9**	0.74	−0.5	1.24	41.6**	0.92	0.5	1.48

Note: * and ** denote significance at 5% and 1% levels. $W_{m,N}^{AF}$ and d_m^{AF} are respectively the BDS and the dimension tests after randomization.

Table 3.5 Maximal Liapunov exponents

	$m=1$	$m=2$	$m=4$
Uniform i.i.d.	3.40	1.41	0.77
Tent map	2.93	0.91	0.36
Rossler map	0.67	0.06	0.09
Industrial prod.	2.68	0.75	0.33
Transp. equip.	1.71	0.60	0.36
Industrial mach.	1.75	0.64	0.28
Electrical mach.	1.81	0.49	0.26
High-tech	1.59	0.46	0.21
Employment	1.55	0.67	0.30
Hourly earnings	1.81	0.70	0.39
Consumer price index	1.88	0.93	0.45

The presence of structures different from those typical of an i.i.d. process has been pointed out by the recurrence plots of all the time series. If we compare Figures 3.9–3.16 with Figure 3.4 (Figure 3.4 is typical of an unstructured random process), we can see clearly the existence of structures (repetitive continuous lines over time) in the distances (represented by the intensity of grey) between the embedded vectors (represented by each single point in the coordinates).[31]

The application of the BDS test gives us further information about the existence of determinism in time series. Applying the BDS test to all the time series at our disposal, we are not able to accept the null i.i.d. hypothesis. All the series are characterized by high values of the BDS statistic well beyond their respective critical values (column $W_{m,N}$, Table 3.4). The dimension test,[32] based, as the BDS test, on the calculus of the correlation dimension, allows us in some cases to measure the dimension of the chaotic attractor that characterizes the time series. Without going into the details, the dimension test is based on the fact that a truly stochastic process is characterized by the growth of the correlation dimension with the increase of the embedding dimension, while a truly chaotic process is characterized by the correlation dimension tending to settle to a constant value when the embedding dimension increases (Hommes 1998). This constant value represents the dimension of the chaotic attractor. In all the series we have analyzed the correlation dimension (column d_m in Table 3.4) grows less than proportionally with respect to m, but in many cases we cannot detect a clear tendency of the correlation dimension to settle clearly to a constant value. For all the time series we have analyzed, the BDS test suggests that the time series contains a deterministic structure, but it is not possible to quantify,

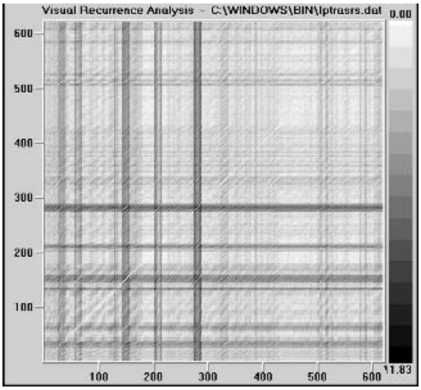


Figure 3.10 *Recurrence plot, residuals, transportation equipment*

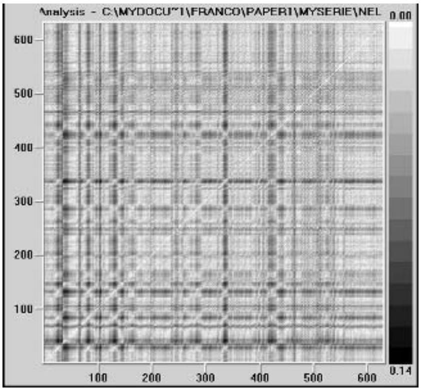


Figure 3.11 *Recurrence plot, residuals, industrial machinery*

via the dimension test, the dimension of the underlying attractor of the time series.[33]

To check further our results we have randomly ordered the real time series, applied BDS and calculated the dimension correlation of the shuffled time series to see whether temporal linkages were relevant. In all the cases the values of the BDS and the dimension tests of the shuffled time series were notably different. We could not reject the null hypothesis of the BDS test for all the shuffled time series (column $W_{m,N}^{AF}$ Table 3.4) and the correlation dimension also was also higher (column d_m^{AF}, Table 3.4) with respect to the original time series. This is confirmation that temporal link-

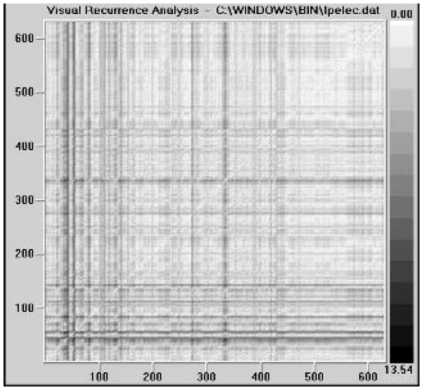


Figure 3.12 Recurrence plot, residuals, electrical machinery

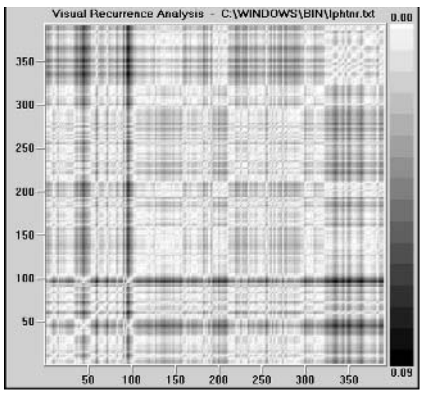


Figure 3.13 Recurrence plot, residuals, high technology

ages between residuals are really important and therefore a mere probabilistic hypothesis on the residuals of macroeconomic time series does not have empirical grounds.

7. CONCLUDING REMARKS

We have first shown the theoretical possibility (Sections 4 and 5) and then the empirical evidence (Section 6) that in the serially uncorrelated residuals there are non-linear signals which, in the models with a deterministic

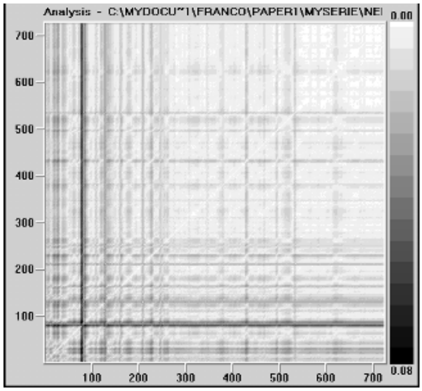


Figure 3.14 *Recurrence plot, residuals, employment*

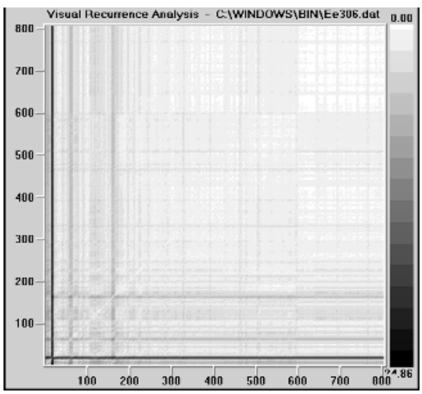


Figure 3.15 *Recurrence plot, residuals, hourly earnings*

(linear or broken) or stochastic trend, are assumed to be i.i.d., like white noise. The approach that we put forward is to separate the stochastic component (that is indeed present in the residuals) from the deterministic component and study these two components separately. To be successful in this task we need a data filter based on the concepts of non-linear dynamics. In this chapter we have limited our analysis to the detection of the existence of clear non-linearities in the residuals of macroeconomic time series. We have detected non-linearities in all the time series we analyzed. All the time series we have considered are thus characterized by determinism, notwithstanding all the series (except employment) are non-stationary and residu-

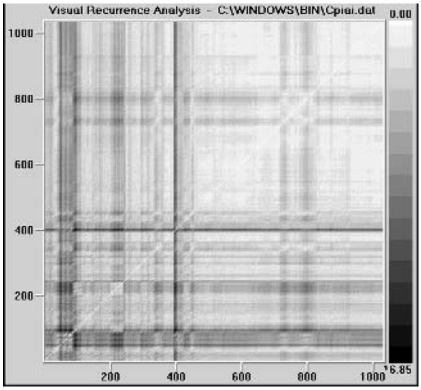


Figure 3.16 Recurrence plot, residuals, consumer price index

als are serially uncorrelated. If all this is true, in the short run, we may make better predictors than simple autoregressive models.

The problem of distinguishing between the two alternative hypothesis, deterministic trend or stochastic trend, was at the core of unit root and broken trend literature (Section 2), but for us it was not the most important issue. Our aim was to detect non-linear structures in those components that linear stochastic models have assumed as exogenous factors. In so far as in linear stochastic models noise plays the relevant role to make 'non-stationary' basically stationary processes, it was for us of primary importance, from the theoretical point of view, to check whether a component of what has been so far assumed as noise might have an endogenous explanation. If this is the case, as confirmed in Section 6, economic variables may not follow a stationary path even in the absence of external shocks and the observed non-stationarity may be the consequence of complex relations between the economic variables.

NOTES

1. Nelson and Plosser have analyzed 14 macroeconomic time series for the USA (with starting date between 1860 and 1909 and with final date 1970). Among these there are real GNP, nominal GNP, industrial production, employment, the unemployment rate, the consumer index rate, nominal wages and real wages.
2. In the classical econometric works, time series were considered stationary along a deterministic trend, that is variables are a linear function of time:

$$x_t = \beta t + \alpha + \varepsilon_t$$

with ε_t i.i.d., α and β parameters, t time and x_t a random variable x observed at time t.

In this case the time series of the variable x is stationary along a time trend and each ε_t has only temporary effects. The short-run component may be insulated, regressing x_t against time and assuming the regression line as the abscissa. This procedure was approximately the one that was used in the 1970s to analyze short-run cycles.

3. In *unit root* processes, time series are not stationary and follow a *random walk*, such as: $x_t = \rho x_{t-1} + \varepsilon_t$ with ε_t i.i.d. and $\rho = 1$. This process is called '*unit*' *root* because x_{t-1} is multiplied by a parameter equal to one (or close to one). It is a '*root*' because one is the root of a characteristic equation (see Enders 1995, p. 25). Each ε_t has persistent effects since, as we can see, each fluctuation will not be reabsorbed in the future: $x_t = x_{t-1} + \varepsilon_t = x_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \ldots = \varepsilon_0 + \varepsilon_1 + \ldots + \varepsilon_{t-1} + \varepsilon_t$. The *signal* x_t is therefore generated by the past and present noise ε. Since noise is an i.i.d. and exogenous variable, we conclude that the variable x_t depends entirely on a variable which we don't know anything about.
4. This result also seems not to depend on the frequency of observation: Wells (1997) and Osborn et al. (1999) have found similar results using both quarterly and monthly data.
5. Except Banerjee et al. (1992), Bresson and Celimene (1995), Dolado and Lopez (1996).
6. The consumer price index and nominal wages for instance were found to follow a random walk.
7. Where the consumption decisions are based on a well-behaved utility function

$$U = \sum_{t=0}^{\infty} \beta^t u(C_t, L_t) \text{ with } \beta < 1,$$

where L_t is the leisure at time t, u the utility. ∞ indicates that the individual is the infinitely lived representative.
8. A short but detailed description of all the methods used in this chapter can be found in Bevilacqua (2001).
9. We exclude the possibility of analyzing any time series of GDP and GNP because of the dearth of data, since these time series are at most quarterly.
10. Links to the files concerning monthly seasonally adjusted and in real terms for industry productions were found at: http://www.bog.frb.fed.us/releases/G17/download2.htm

    Indices of industrial production go back to 1919 and the respective base year is 1992.

    A table showing the historical consumer price index for all urban consumers beginning in 1913 was available from the Bureau of Labor and Statistics at: ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt.

    This table refers to all urban consumers, with 1982 as the base year.

    The seasonally adjusted 'hourly wages' time series in this chapter refers to manufacturing industry with data of the type 'average hourly earnings of production workers'.
11. As calculated by E. Kononov (1999), VRA 4.2 program.
12. See Section 5.3.1: the case of the *tent map*.
13. Such as, for instance, the Rossler map in Section 5.3.2.
14. This step is also sometimes called 'shuffle diagnostic' (see Lorentz 1989) via 'surrogate time series' (Kantz and Schreiber 1997). A 'surrogate' time series is essentially the shuffle of the original time series preserving all the linear properties of the time series like frequencies, amplitudes and eventual linear autocorrelations. We have derived the surrogate time series for all the economic time series we have analyzed, but we called them the more general and less specialistic term of 'shuffled time series'.
15. Note that ε in all our experiments is distributed as a uniform distribution. Similar results can be obtained using other distributions, such as the normal. However, what is important is that ε is i.i.d. whatever its distribution. We have chosen to use the uniform distribution because in Section 5.3 we show the deterministic case of the tent map which produces ε that are uniformly distributed.
16. We could obviously plot residuals $\varepsilon(t)$ against the residuals of any preceding period, for example $\varepsilon(t-4)$. Knowing *ex ante* that ε is the result of a random number generator, the $\varepsilon(t) - \varepsilon(t-4)$ plot is qualitatively equivalent to the $\varepsilon(t) - \varepsilon(t-1)$ plot.

17. The embedded vectors are simply defined as:

$$\mathbf{x}_i = \{x_i - (m-1), x_i - (m-2), ..., x_i\}$$

where x_i is the observed value at a certain point at time and m is called embedding dimension.

For example, suppose we have a series of ten observed values of a certain variable x:

$$x = \{8, 5, 6, 9, 4, 4, 1, 7, 3, 2, 7\},$$

we obtain the following embedded vectors:

$$\mathbf{x}_2 = \{x_{2-(2-1)}, x_{2-(2-2)}\} = \{x_1, x_2\} = \{8,5\}$$
$$\mathbf{x}_3 = \{x_{3-(2-1)}, x_{3-(2-2)}\} = \{x_2, x_3\} = \{5,6\}$$
$$\mathbf{x}_4 = \{x_{4-(2-1)}, x_{4-(2-2)}\} = \{x_3, x_4\} = \{6, 9\}$$
$$\cdots$$
$$\mathbf{x}_{10} = \{x_{10-(2-1)}, x_{10-(2-2)}\} = \{x_9, x_{10}\} = \{3,7\}$$
for $m = 2$,
and $\mathbf{x} = \{x_2, x_3, x_4, \ldots, x_{10},\}$ is the embedded time series for $m = 2$.

The embedded time series are of great importance in non-linear dynamics because, thanks to them, as has been shown by Takens (1981), we may uncover some properties such as the correlation dimension of an unknown underlying motion law that generated the time series itself from the observed values of the process.

18. That is, between vectors $\mathbf{x}_i$.
19. Because of the finite approximation of the program we used, we could not obtain more than 50 observations. Consequently we have added a very small ratio of white noise to each ε_t so that the system does not repeat itself even in the long run. We have added 0.000001 * U(0.5, 1) noise.
20. See Bevilacqua (2001) or the original work by Brock et al. (1991) for the size and power of the BDS test.
21. For a detailed description of the Rossler process see Lorentz (1989) or Gandolfo (1997).
22. A generally accepted result is that the GDP time series, as pointed out by the vast literature on unit roots and co-integration, is characterized by a stochastic trend, but it cannot be reliably tested with the non-linear numerical tools because of a paucity of observations. Hence we cannot ascertain whether the GDP is really characterized by a non-linear dynamics.
23. Since some time series were autocorrelated in the residuals, we have used for all the real time series the 'augmented' form of the Dickey–Fuller test including more lags, trend and intercept. The number of lags we have considered is the minimal that allows us to obtain uncorrelated residuals. See Harris (1995) for more details.
24. All the sector time series we have considered are in terms of value.
25. Obtained setting $m = 5$.
26. Similar results were also obtained by adding a small percentage of noise (5 percent of the variance). We added noise to the time series simply because, when the non-linear structure is well defined, adding a small stochastic component should not change significantly the result of the test. Even if there were small i.i.d. measure errors, these should not call in question the obtained results.
27. Those that are the most important with respect to the value added.
28. However, for transportation equipment production and industrial machinery production we are not able to reject the null hypothesis at the 1 percent significance level.
29. See the estimated equations within Table 3.4.
30. It is worth mentioning that in Section 6.1 we found a maximal Liapunov exponent for industrial production close to zero, indicating the presence of cycles.
31. The presence of continuous lines in the recurrence plots indicates that the embedded vectors represented by each point keep approximately the same distance with respect to

all the vectors that belong to the continuous line. In a normal i.i.d. process, each vector is randomly distant from any other vector and the probability that nearby vectors have similar distances is very low. Thus in a normal i.i.d. process we should not notice any continuous line in the recurrence plots.

32. Note that the 'dimension test', contrary to the BDS test, is not really a statistical test since critical values are not specified. It's a numerical tool that suggests the existence of a deterministic dynamics when the calculated correlation dimension tends towards a fixed value when the embedding dimension grows.
33. This phenomenon may be due to the presence of a stochastic component in the time series. It should therefore be important to filter our data in order to separately analyze the deterministic component and to quantify the dimension of the chaotic attractor. The future application of filters that allow us to reduce and, it is hoped, remove the stochastic component may allow us to detect the dimension of chaos for all the real time series for which we have already uncovered the presence of chaos.

REFERENCES

Alba, J.D. and Papell, D.H. (1995), 'Trend breaks and the unit root hypothesis for newly industrializing and newly exporting countries', *Review of International Economics*, **3**, 264–74.

Al Bazai, H.S. (1998), 'The Saudi stock market and the monetary policy', *Journal of the Social Sciences*, **26**, 91–106.

Banerjee, A., Lumsdaine, R.L. and Stock, J.H. (1992), 'Recursive and sequential tests of the unit root and trend break hypothesis: theory and international evidence', *Journal of Business and Economic Statistics*, **10**, 271–87.

Ben-David, D. and Papell, D.H. (1994), 'The great wars, the great crash, and the unit root hypothesis: some new evidence about an old stylized fact', CEPR, DP. 965.

Bevilacqua, F. (2001), 'Nonlinear dynamics in US macroeconomic time series', Merit-Infonomics RM 2001–035.

Bohl M.T. (1998), 'Kovergenz westdeutscher Regionen? Neue empirische Ergebnisse auf der Basis von Panel-Einheitswurzeltests', *Konjunkturpolitik*, **44**, 82–99.

Boswijk, H.P. (1996), 'Unit roots and cointegration', lecture notes, Amsterdam University.

Bresson, G. and Celimene, F. (1995), 'Hysteresis or persistence of the unemployment rate: an econometric test in several Carribean countries', *Revue d'Economie Politique*, **105**, 965–98.

Brock, W.A. and Dechert, W.D. (1988), 'Theorems on distinguishing deterministic from random systems', in W. Barnett et al., *Dynamic Econometric Modelling*, Cambridge: Cambridge University Press.

Brock, W.A., Dechert, W.D. and Scheinkman, J.A. (1987), 'A test for independence based on the correlation integral', University of Wisconsin–Madison.

Brock, W.A., Hsieh, D.A. and LeBaron, B. (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, Cambridge, MA: MIT Press.

Capitelli, R. and Scjlegel, A. (1991), 'Ist die Schweiz eine Zinsinsel? Eine multivariate Untersuchung über den langfristigen internationalen Realzinsausgleich', *Swiss Journal of Economic and Statistics*, **127**, 647–64.

Caselli, G.P. and Marinelli, L. (1994), 'Italian GNP growth 1890–1992: a unit root or segmented trend representation?', *Economic Notes*, **23**, 53–73.

Chavas, J.P. and Holt, M.T. (1991), 'On nonlinear dynamics: the case of the Pork Cycle', *American Journal of Agricultural Economics*, **73**, 819–28.
Chavas, J.P. and Holt, M.T. (1993), 'Market instability and nonlinear dynamics', *American Journal of Agricultural Economics*, **75**, 113–20.
Cheung, Y.W. and Chinn, M.D. (1996), 'Further investigation of uncertain unit root in GNP', NBER, 206.
Choi, I. and Yu, B. (1997), 'A general framework for testing I(m) against I(m+k)', *Journal of Economic Theory and Econometrics*, **3**, 103–38.
Coakley, J., Kulasi, F. and Smith, R. (1995), 'Current account solvency and the Feldstein–Horioka puzzle', London: Birkbeck College Discussion Paper.
Conover, W.J. (1971), *Practical Nonparametric Statistics*, New York: Wiley.
Coorey, S. (1991), 'The determinants of U.S. real interest rates in the long run', IMF WP 118.
Crosby, M. (1998), 'A note on the Australian business cycle', *Economic Analysis and Policy*, **28**, 103–8.
Culver, S.E. and Papell, D.H. (1995), 'Real exchange rates under the Gold Standard: can they be explained by the trend break model?', *Journal of International Money and Finance*, **14**, 539–48.
Cushing, M.J. and McGarvey, M.G. (1996), 'The persistence of shocks to macroeconomic time series: some evidence from economic theory', *Journal of Business and Economic Statistics*, **14**, 179–87.
Dechert, W.D. (1994), 'The correlation integral and the independence of Gaussian and related processes', SSRI W.P. 9412, University of Wisconsin.
Dickey, D.A. and Fuller, W.A. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, **74**, 427–31.
Diebold, F.X. and Rudebusch, G.D. (1989), 'Long memory and persistence in aggregate output', *Journal of Monetary Economics*, **24**, 189–209.
Dolado, J.J. and Lopez, S.J.D. (1996), 'Hysteresis and economic fluctuations (Spain, 1970–94)', Centre for Economic Policy Research, Discussion Paper 1334.
Dolmas, J., Raj, B. and Slottje, D.J. (1999), 'A peek through the structural break window', *Economic Inquiry*, **37**, 226–41.
Duck, N.W. (1992), 'Evidence on breaking trend functions from nine countries', University of Bristol WP 341.
Durlauf, S.N. (1993), 'Time series properties of aggregate output fluctuations', *Journal of Econometrics*, **56**, 39–56.
Eckmann, J.P., Kamphorst, O. and Ruelle, D. (1987), 'Recurrence plots of dynamical systems', *Europhysic Letters*, **4**, 973–9.
Enders, W. (1995), *Applied Econometric Time Series*, New York: Wiley.
Fleissing, A.R. and Strauss, J. (1997), 'Unit root tests on real wage Panel Data for the G7', *Economic Letters*, **56**, 149–55.
Frank, M. and Stengos, T. (1988), 'International Chaos?', *European Economic Review*, **32**, 1569–84.
Franses, P.H. and Kleibergen, F. (1996), 'Unit roots in the Nelson Plosser data: do they matter for forecasting?', *International Journal of Forecasting*, **12**, 283–88.
Fuller, W.A. (1999), *Introduction to Time Series*, New York: Wiley.
Fung, H.G. and Lo, W.C. (1992), 'Deviations from purchasing power parity', *Financial Review*, **27**, 553–70.
Gallegati, M. (1996), 'Testing output through its supply components', *Economic Notes*, **25**, 249–60.

Gamber, E.N. and Sorensen, R.L. (1994), 'Are net discount rates stationary?', *Journal of Risk and Insurance*, **61**, 503–12.

Gandolfo, G., 1997, *Economic Dynamics*, New York and Heidelberg: Springer Verlag.

Gil, A.L.A. and Robinson, P.M. (1997), 'Testing of unit roots and other non-stationary hypothesis in macroeconomic time series', *Journal of Econometrics*, **80**, 241–68.

Gokey, T.C. (1990), 'Stationarity of nominal interest rates, inflation and real interest rates', Oxford Applied Economics DP 105.

de Haan, J. and Zelhorst, D. (1994), 'Testing for stationarity of output components: some results for Italy', *Economic Notes*, **23**, 402–9.

Hansen G. (1991), 'Hysteresis and unemployment', *Jahrbücher für Nationalökonomie und Statistik*, **208**, 272–98.

Harris, R. (1995), *Using Cointegration Analysis in Econometric Modelling*, London: Prentice Hall.

Haslag, J., Nieswiadomy, M. and Slottje, D.J. (1994), 'Are net discount rates stationary? Some empirical evidence', *Journal of Risk and Insurance*, **61**, 513–18.

Hommes, C.H. (1998), 'Nonlinear economic dynamics', lecture notes, NAKE Utrecht.

Hsieh, D.A. (1991), 'Chaos and nonlinear dynamics: application to financial markets', *Journal of Finance*, **46**, 1839–77.

Hylleberg, S. and Engle, R.F. (1996), 'Common seasonal features: global unemployment', Aarhus Department of Economics WP 13.

Kantz, H. and Schreiber, T. (1997), *Nonlinear Time Series Analysis*, Cambridge: Cambridge University Press.

Kennel, M.B., Brown, R. and Abarbanel, H.D.I. (1992), 'Determining embedding dimension for phase-space reconstruction using a geometrical construction', *Physics Review*, **45**, 3403–11.

King, R.G., Plosser, C.I. and Rebelo, S.T. (1988a), 'Production, growth and business cycles I: the basic neoclassical model', *Journal of Monetary Economics*, **21**, 195–232.

King, R.G., Plosser, C.I. and Rebelo, S.T. (1988b), 'Production, growth and business cycles II: new directions', *Journal of Monetary Economics*, **21**, 309–41.

Kononov, E. (1999), 'Visual Recurrence Analysis 4.2 (software program)', http: //home. netcom.com/~eugenek/.

Krol, R. (1992), 'Trends, random walks and persistence: an empirical study of disaggregated U.S. industrial production', *Review of Economics and Statistics*, **74**, 154–9.

Lee, H S. and Siklos, P.L. (1991), 'Unit roots and seasonal unit roots in macroeconomic time series: Canadian evidence', *Economic Letters*, **35**, 273–77.

Lee, J. (1996), 'Testing for a unit root in time series with trend breaks', *Journal of Macroeconomics*, **18**, 503–19.

Leslie, D., Pu, Y. and Wharton, A. (1995), 'Hysteresis versus persistence in unemployment: a sceptical note on unit root tests', *Labour*, **9**, 507-23.

Leybourne, S.J. and McCabe, B.P.M. (1999), 'Modified stationary tests with data dependent model selection rules', *Journal of Business and Economic Statistics*, **17**, 264–70.

Leybourne, S.J., McCabe, B.P.M. and Tremayne, A.R. (1996), 'Can economic time series be differenced to stationarity?', *Journal of Business and Economic Statistics*, **14**, 435–46.

Li, H. (1995), 'A reexamination of the Nelson–Plosser data set using recursive and sequential tests', *Empirical Economics*, **20**, 501–18.

Linden, M. (1992), 'Stochastic and deterministic trends in Finnish macroeconomic time series', *Finnish Economic Papers*, **5**, 110–16.

Lorentz, H.W. (1989), *Lecture Notes in Economics and Mathematical Systems*, Berlin: Springer Verlag.

Lucas, R. (1972), 'Expectations and the neutrality of money', *Journal of Economic Theory*, **4**, 103–24.

Lucas, R. (1977), 'Understanding Business Cycles', in K. Brunner and A.H. Meltzer (eds), *Stabilization of the domestic and international economy*, Carnegie–Rochester Conference Series on Public Policy, **5**, 7–29.

Lucas, R. (1980), 'Methods and problems in Business Cycle Theory', *Journal of Money, Credit, and Banking*, **12**, 696–715.

Lumsdaine, R.L. and Papell, D.H. (1997), 'Multiple trend breaks and the unit root hypothesis', *Review of Economics and Statistics*, **79**, 212–18.

MacDonald, R. (1996), 'Panel unit root tests and real exchange rates', *Economic Letters*, **50**, 7–11.

Maddala, G.S. and Kim, I.M. (1998), *Unit roots, cointegration and structural change*, Cambridge: Cambridge University Press.

McCoskey, S.K. and Selden, T.M. (1998), 'Health care expenditures and GDP: panel data unit root test results', *Journal of Health Economics*, **17**, 369–76.

McDougall, R.S. (1995), 'The seasonal unit root structure in New Zealand macroeconomic variables', *Applied Economics*, **27**, 817–27.

Mills, T. (1992), 'How robust is the finding that innovations to UK output are persistent?', *Scottish Journal of Political Economy*, **39**, 154–66.

Mocan, H.N. (1994), 'Is there a unit root in U.S. real GNP? A re-assessment', *Economic Letters*, **45**, 23–31.

Moosa, I.A. and Bhatti, R.H. (1996), 'Does real interest parity hold? Empirical evidence from Asia', *Keio Economic Studies*, **33**, 63–70.

Nelson, C.R. and Plosser, C.I. (1982), 'Trends and random walks in macroeconomic time series: some evidence and implications', *Journal of Monetary Economics*, **10**, 139–62.

Nunes, L.C., Newbold, P. and Kuan, C.M. (1997), 'Testing for unit roots with breaks: evidence on the great crash and the unit root hypothesis reconsidered', *Oxford Bulletin of Economics and Statistics*, **59**, 435–48.

Oppenheim, A.V. and Schafer, R.W. (1989), *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall.

Osborn, D.R., Heravi, S. and Birchenhall, C.R. (1999), 'Seasonal unit roots and forecasts of two digit European industrial production', *International Journal of Forecasting*, **15**, 27–47.

Parikh, A. (1994), 'Tests of real interest parity in international currency markets', *Journal of Economics*, **59**, 167–91.

Perron, P. (1989), 'The great crash, the oil price shock and the unit root hypothesis', *Econometrica*, 1361–401.

Prescott, E. (1998), 'Notes on business cycle theory: methods and problems', ISER draft, Siena.

Rahman, M. and Mustafa, M. (1997), 'Dynamics of real exports and real economic growths in 13 selected Asian countries', *Journal of Economic Development*, **22**, 81–95.

Raj, B. (1992), 'International evidence on persistence in output in the presence of episodic change', *Journal of Applied Econometrics*, **7**, 281–93.

Raj, B. and Slottje, D.J. (1994), 'Are trend behavior of alternative income inequality measures in the United States from 1947–1990 and the structural break', *Journal of Business and Economic Statistics*, **12**, 479–87.
Rappoport, P. and Reichlin, L. (1986), 'On broken trends, random walks and nonstationary cycles', in M. Di-Matteo, R.M. Goodwin and A. Vercelli (eds), *Technological and social factors in long term fluctuations*, Proceedings of ISER Workshop held in Siena, Italy, 16–18 December 1986. Lecture Notes in Economics and Mathematical Systems series, no. 321, Springer, 1989, 305–31.
Rappoport, P. and Reichlin, L. (1988), 'Segmented trends and nonstationary time series', Federal Reserve Bank of New York.
Rappoport, P. and Reichlin, L. (1989), 'Segmented trends and nonstationary time series', *Economic Journal*, **99**, supplement, 168–77.
Rudebusch, G.D. (1990), 'Trends and random walks in macroeconomic time series: a re-examination', Board of Governors of the Federal Reserve System Finance and Economics Discussion Series, 139.
Rudebusch, G.D. (1992), 'Trends and random walks in macroeconomic time series: a re-examination', *International Economic Review*, **33**, 661–80.
Scheinkman, J.A. (1990), 'Nonlinearities in Economic Dynamics', *Economic Journal,* **100**, Supplement, 33–48.
Serletis, A. (1994), 'Breaking trend functions in the velocity of money: evidence from the United States and Canada', *North American Journal of Economics and Finance*, **5**, 201–8.
Serletis, A. and Zimonopoulos, G. (1997), 'Breaking trend functions in real exchange rates: evidence from seventeen OECD countries', *Journal of Macroeconomics*, **19**, 781–802.
Simkins, S.P. (1994), 'Business cycles, trends, and random walks in macroeconomic time series', *Southern Economic Journal*, **60**, 977–88.
Soejima, Y. (1995), 'A unit root test with structural change for Japanese macroeconomic variables', *Monetary and Economic Studies*, **13**, 53–68.
Solow, R. (1956), 'A contribution to the theory of economic growth', *Quarterly Journal of Economics*, **70**, 65–94.
Song, F.M. and Wu, Y. (1997), 'Hysteresis in unemployment: evidence from 48 U.S. States', *Economic Inquiry*, **35**, 235–43.
Song, F.M. and Wu, Y. (1998), 'Hysteresis in unemployment: evidence from OECD countries', *Quarterly Review of Economics and Finance*, **38**, 181–92.
Sosa, E.W. (1997), 'Testing for unit roots and trend breaks in Argentine real GDP', *Economica*, **43**, 123–42.
Swan, T.W. (1956), 'Economic growth and capital accumulation', *Economic Record*, **32**, 334–61.
Takens, F. (1981), *Detecting strange attractors in turbulence*, Lecture Notes in Mathematics, Springer.
Tong, H. (1990), *Non-linear Time Series: a Dynamical System Approach*, Oxford: Clarendon Press.
Vastano, J.A. and Wolf, A. (1986), 'Intermediate Length Scale Effects in Liapunov Exponent Estimation', Department of Physics and Center for Nonlinear Dynamics, University of Texas.
Weliwita, A. (1998), 'Cointegration tests and the long run purchasing power parity: examination of six currencies in Asia', *Journal of Economic Development*, **21**, 103–15.
Wells, J.M. (1997), 'Modelling seasonal patterns and long run trends in U.S. time series', *International Journal of Forecasting*, **13**, pp. 407–20.

Wu, Y. (1996), 'Are real exchange rates nonstationary? Evidence from a panel data test', *Journal of Money, Credit and Banking*, **28**, 54–63.
Wu J.L. and Chen S.L., 1995, 'Deterministic or stochastic trends: the empirical evidence from Taiwan', *Taiwan Economic Review*, **23**, 223–37.
Wu, J.L. and Chen, S.L. (1998), 'A re-examination of real interest rate parity', *Canadian Journal of Economics*, **31**, 837–51.
Wu, P. and Crato, N. (1995), 'New tests for stationarity and parity reversion: evidence on New Zealand real exchange rates', *Empirical Economics*, **20**, 599–613.
Wu, Y. and Zhang, H. (1996), 'Mean reversion in interest rates: new evidence from a panel of OECD countries', *Journal of Money, Credit and Banking*, **28**, 604–21.
Zelhorst D. and de Haan J. (1993), 'The nonstationarity of German aggregate output', *Jahrbücher für Nationalökonomie und Statistik*, **212**, 410–18.
Zivot, E. and Andrews, D. (1990), 'Further evidence on the great crash, the oil price shock, and the unit root hypothesis', Yale Cowles Foundation DP 944.
Zivot, E. and Andrews, D. (1992), 'Further evidence on the great crash, the oil price shock, and the unit root hypothesis', *Journal of Business and Economic Statistics*, **10**, 251–70.

4. The use of genetic programming in evolutionary economics

Bernd Ebersberger and Andreas Pyka

1. INTRODUCTION

Evolutionary economics has recently been labeled as 'among the most hopeful, and . . . most fruitful, developments in economics' (Blaug 1998, p. 31). This favorable assessment is based on the new modeling approaches that have largely replaced the 'social mathematics' (Blaug 1998, p. 11) that mainstream economic analysis tends to favor.

These new modeling approaches in evolutionary economics reflect its demand for a new type of modeling (cf. Boulding 1991, p. 51). Among others, they include techniques such as cellular automata,[1] neural networks,[2] master equation approaches,[3] and genetic algorithms and genetic programming.[4]

In this chapter we will build on the successful introduction of those new modeling techniques and genetic programming in particular which have been applied to various types of modeling tasks in evolutionary economics. We suggest in this chapter that genetic programming can also be an instrument in the economists' toolbox to improve existing models or to support the generation of new economic models that relate to empirical observations. Hence, we claim that genetic programming can also be used for empirical analysis.

We do not confine the term *empirical analysis* to the mere testing of hypotheses. Rather, we maintain a broad concept of the term *empirical analysis* including any step in the research process that confronts economic theory with observed data.

In this chapter we want to introduce a rationale for the use of genetic programming in empirical analysis. In doing so we want to start from the very foundations of evolutionary economics and base our arguments on the bounded rationality of the agents inhabiting the models in evolutionary economics.

Hence this chapter starts from the bounded rational agent in Section 2. In Section 2.2 we argue that the learning of bounded rational agents can be modeled by genetic programming. In Section 3.1 we set up a thought experiment where an economist is assumed to be boundedly rational. We argue that genetic programming can be used to simulate the learning of a boundedly rational economist. Such learning is essentially obtaining improved models of economic reality. In a subsequent step in Section 3.2 we argue that the insights from the thought experiment can be applied independently of the simulation task. Genetic programming can be used as a stand-alone tool to improve economic models. An application of how genetic programming can be applied to improve a given model is discussed in Section 4. Section 5 concludes.

2. GENETIC PROGRAMMING TO MODEL THE LEARNING OF BOUNDED RATIONAL AGENTS

In this section we discuss briefly how the learning of bounded rational agents can be modeled. In the introduction above we used the term 'modeling' without explaining exactly what it means. Hence, before we discuss the modeling of the bounded rational agent, we must elaborate on what is meant by 'modeling'.

2.1 What is Modeling?

In a very general statement we can say that a model describes reality. Hence we can think of 'modeling' as the process of *mapping reality into a language*.[5] To be able to formulate a model as an image of reality one has to accept certain simplifications of the complex world (for example Williams 1999, p. 15). A model can only describe a *small fraction of the reality* and the description of a fraction of the reality *requires abstraction*. No feature of the real-world phenomenon can be mapped into the language description (cf. Casti 1992, p. 2). Here, abstraction means that the model *contains relevant features* of reality. Irrelevant details of reality, however, are disregarded in the modeling process (see Baumol and Blinder 1999, p. 9).

Although the modeling process has to be regarded as the description of reality, it is not carried out for its own sake. The use of a model proceeds beyond pure description, as do models' purposes. To use a model we need a formulation of it that *allows for manipulation*, because we must manipulate the model to make inferences about reality (Zwicker 1981, p. 19).

Summing up the discussion above, a model is a simplified, abstract, manipulable (language-)representation of a fraction of reality.

2.2 The Learning of Bounded Rational Agents

Evolutionary economics presents an ontological challenge to the current mainstream economics as it regards the economic system as open (cf. Hodgson 1998, p. 165; Foss 1994, p. 35). In an open environment agents cannot be considered as having perfect information and perfect foresight. To make things even worse, agents are supposed to be endowed with limited computational power. Restrictions on other resources such as memory or time pose additional limitations on the rationality of agents (Dosi and Egidi 1991; Egidi et al. 1992). Evolutionary economics maintains bounded rationality of agents.

We assume that agents carry internal descriptions of their real environment. Due to the mental capabilities of the actors, these descriptions are simplifications of reality. They are manipulable, as agents can use them for various purposes such as predicting the outcome of certain actions and so on. This suggests that we can think of agents as carrying mental models of their environment. When we are talking about the agent's *representation of the environment* or his *mental model*, we are referring to the agent's perception of reality, that is, how he or she envisages his or her environment.

The lack of perfect foresight implies that actors do not possess mental models of their environment and the economic system in particular that amount to a perfect description. The fact that the mental models are manipulable suggests that actors can revise and update them. This process of changing the mental representation of reality can be labeled 'learning'. Hence learning is about the agents' creation of mental models. According to the discussion above, the formulation of mental models can be called 'modeling'. Hence we suppose that learning processes consist of agents mentally modeling their environment. Having said this, we can think of learning as modeling.

If we now model the agent and his learning process, we have to model learning (of the modeled agent) as modeling (Moss and Edmonds, 1998).

Modeling the learning of a bounded rational agent rests on the assumption that the agent maintains a number of mental models (Edmonds 1999a, p. 306). The agent also has a number of observations that can be used to determine which model best describes reality. While this mental model is the ruling one, the other models are not discarded. Rather, the agent tries constantly to challenge the currently ruling model. To do so he constantly rearranges and combines all the maintained mental models. Once there is a model that describes reality better than the currently ruling model, it is replaced by the better model.

This selection and the combining and rearranging process can be modeled using genetic programming (cf. Edmonds 1999a, 1999b; Moss and

Edmonds 1998). The genetic programming algorithm (GP algorithm) belongs to the class of evolutionary algorithms. Genetic programming was introduced by Koza (1992 and 1994), and gave rise to an enormous number of publications.[6] For an introduction to genetic programming see Banzhaf et al. (1998) or Koza (1992).

The genetic programming model of the bounded rational agent incorporates the idea that agents carry a large number of mental descriptions of reality. The mental models of the agents in genetic programming are language constructs generated from a predefined set of language components.[7] The set of mental models in the genetic programming model is denoted population.

In the context of genetic programming the criterion used determine how well a mental model describes reality is referred to as the fitness function. The modification and the structural change in the mental models are performed by genetic operators such as mutation, or cross-over. Mutation performs a random change in the structure and the content of a mental model, whereas cross-over combines the structure and content of two or more mental models. It should be mentioned here that the set-up of the genetic programming routine has to be such that the genetic operators yield syntactically valid models. In the original set-up this is achieved by imposing closure on the elements used (Koza 1992, p. 81; Banzhaf et al. 1998, p. 112). Once there is a mental model in the population with a better fitness compared to the currently ruling mental model, the current model is replaced.

Edmonds (1999a) shows that genetic programming is appropriate to model the learning of bounded rational agents for various reasons:

1. The population of the programs in genetic programming models the number of mental models carried by the actors.
2. If, *ceteris paribus*, the selection criterion of the agent changes, there are several alternative mental models that can be used instead of the current one.
3. The agents are capable of acting, as there is at least one best mental model in the population of mental models.
4. The one best model does not mean that it is the best or even a good mental model to describe reality. However, relative to the other models maintained it is the best. (*The one-eyed man is king in the valley of the blind.*)
5. Thus modeling the bounded rational agent with genetic programming implicitly means that the agent can be wrong.
6. The fitness function can be implemented so as to relate the observations of the agent to the mental models.

7. As genetic programming allows for modeling the changing structure and changing content of the mental models, it might generate some degree of creativity in developing new mental models.
8. The whole mental capacity of the agent can be described in terms of the whole population of mental models he carries.
9. The limitations of resources, such as memory, computational power can be captured by changing and augmenting the fitness function.
10. Cross-over is the recombination of existing mental models. As cross-over is one means to create new mental models, the generation of new mental models is highly contingent on the existing models. Hence it is path-dependent.

3. MODELING BY MEANS OF GENETIC PROGRAMMING

In this section we want to shed a different light on the use of genetic programming in economics. The previous section showed that genetic programming is used as a model of bounded rational agents. This section draws on the insights gained there and shows that genetic programming can be used as a tool to improve given or generate new models based on available observations.

3.1 Thought Experiment: Modeling an Economist

In the section above we elaborated on modeling a bounded rational agent. At first sight this does not relate to empirical economics at all. However, the main hypothesis of this chapter is that empirical economics can gain from the modeling techniques used in representing the learning of the bounded rational agent. In this section we want to pave the way for the major point of this chapter by using a thought experiment: *imagine we want to model an economist learning about economic reality*.

3.1.1 Some assumptions about economists

To model an economist we have to make certain assumptions about the features of a real economist.

Assumption 1 *Real-world economists are agents trying to learn about reality and economic reality in particular.*

Assumption 2 *In doing so, economists constantly try to improve their perception of reality.*

Assumption 3 *The economists try to formulate their perception of economic reality in models.*

Assumption 4 *Empirical analysis in economics is guided by the desire to improve the understanding of economic reality.*

Assumption 5 *In their endeavor to understand economic reality economists are bounded rational agents:*

1. *Economists do not have perfect information about economic reality.*
2. *Economists are endowed with only limited power to process the information.*
3. *Economists have limited other resources, such as time and memory.*

Based on Assumptions 1 to 5, we conclude that economists are quite similar to the bounded rational agents we discussed above both in their capabilities and their attempt to understand reality on the basis of imperfect information. So our thought experiment of modeling an economist has to be refined as modeling a bounded rational economist.

3.1.2 Modeling a bounded rational economist

The economist starts his endeavor to find a model to describe a given phenomenon with a population of models.[8] He evaluates each of the models on how well they fit the available observations. Based on this evaluation, the economist chooses his current model.

In his endeavor to improve his models the economist comes up with new models. These models might be minor modifications of already existing models, they might be combinations of already existing models. The economist might even consider models that are of totally different structure compared to the already existing models. When trying to improve given models, economists are likely to take good models as a starting point for modification.

Based on how well the models fit the observations, the real-world economist discards the ruling model and replaces it with a new model that fits the observations better. In evaluating how well a model fits reality the economist does not only take his observations into account. He also relates the model to his previous (theoretical) knowledge. A model that contradicts the previous theoretical knowledge is very likely to receive a bad evaluation.

Even though the description of the economist here seems to be quite crude, we think that it captures some of the features of empirical work in economics. Models are constantly checked for how well they fit available observations. Although the discussion above focused on a single economist, the search is not necessarily conducted by one individual only. The search for a model that matches reality better than the currently available models can also be thought of as a collective endeavor carried out by all economists through referencing, criticizing and modifying models taken from literature. Hence learning is both an individual process and a collective process of optimizing the models (Greene 2000, p. 221).

In the following paragraphs we want to sketch how to simulate an economist based on the insights from modeling bounded rational agents. As in modeling a bounded rational agent above, the tool to model an economist will be genetic programming.

Language As we are modeling an economist's learning as modeling, we have to define the language the economic model of reality will be formulated in. The language most often used to formulate economic models is a mathematical language consisting of operators, functions, variables and parameters (or constants). In this case the modeling language a simulated economist can use is given by the primitive function set $F = \{f_1, f_2, f_3, \ldots, f_k\}$ containing the operators and functions and the terminal set $T = \{\tau_1, \tau_2, \tau_3, \ldots, \tau_l\}$. The terminal set contains the variables, the parameters and some arbitrary constants.[9] In a basic set-up the language definition has to satisfy both sufficiency and closure conditions (Koza 1992).

Model evaluation and selection To be able to select a model we have to evaluate how well each fits the available data.

1. *Fitting the observations*: The evaluation is based on the model's fit to the observations that are available to the economist. Various statistical measures of the model fit can be used to do so. In the context of genetic programming, however, the most common fitness measure is based on the sum of squared errors.
2. *Conform with prior knowledge*: Real-world economists not only evaluate the models based on their fitting of the observations, but also on how much they are in accordance with prior theoretical knowledge. One way to include prior knowledge in the fitness function is to derive conditions that a model has to fulfill such as to comply with the prior knowledge. If those conditions are not satisfied by a model it is assigned an unfavorable fitness value.[10]
3. *Selection*: It is an essential feature of genetic programming that better models have a higher probability of being selected for the modification and recombination (cf. Banzhaf et al. 1998, p. 112).

Modification of the mental models Creation of a new set of mental models can be represented by a combination of various genetic operators (Koza

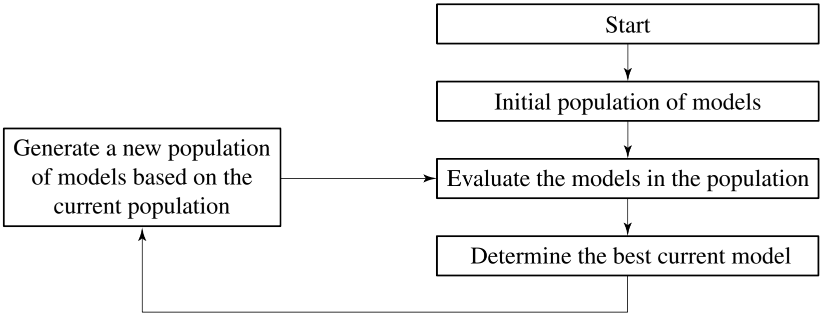


Figure 4.1 Flow chart for the model of a bounded rational economist

1992, p. 83; Banzhaf et al. 1998, p. 125). Modification of the current set of models that involves changes in the content and structure of the models can be represented by a mutation operator. The mutation operator can account for small changes in the content and the structure of the models. It can also account for totally new models. Combination of two mental models or combination of parts of two models can be simulated by the cross-over operator.

The modified models can either replace the models from which they were created or can be used to create a new population of models.

Flow of the learning process Now that we have discussed the main components for modeling a bounded rational economist, we sketch the flow of the learning process. Figure 4.1 shows the flow of the economist's process of improving his mental models.

3.2 Modeling by Means of the GP Algorithm

In the thought experiment in Section 3.1 we argued that in principle it is possible to model an economist's search for improved models using genetic programming. Separating the task of simulating a bounded rational economist from the methodology used to do so enables us to treat genetic programming as a tool to improve models in economics. In this context improvement refers to the models' increased fit to available observations. By carefully designing a genetic programming routine, we can use the simulated agent to improve economic models. Hence genetic programming can be seen as a tool to support modeling.

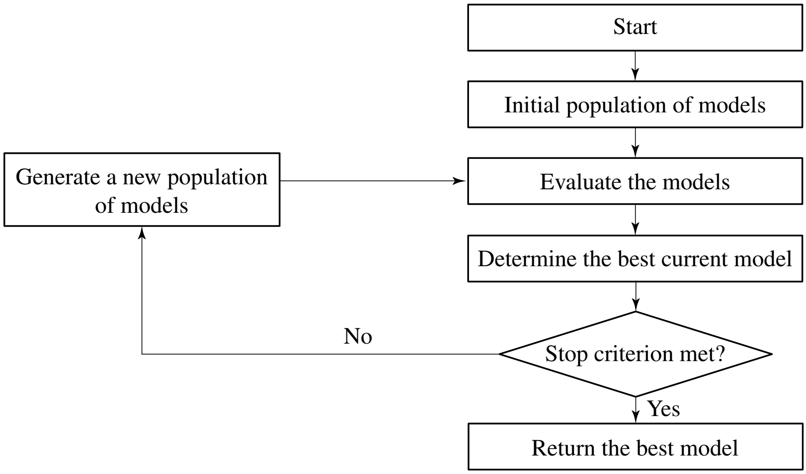


Figure 4.2 Flow chart for the genetic programming routine to be used to improve economic models

3.2.1 The flow of the GP algorithm

Although we can use the basic set-up discussed above, we have to implement minor modifications. We have not defined an end to the learning of the bounded rational agents or to the model-improving economist in the thought experiment. A genetic programming routine theoretically running forever, trying to improve the obtained results even further will not serve any purpose in empirical research. Hence a termination criterion has to be defined in terms of a maximum computing time and in terms of a satisfactory fit of the generated model (see Figure 4.2).

3.2.2 Properties of GP as a modeling support tool

The properties of genetic programming for empirical analysis in economics, particularly in evolutionary economics, are discussed in Ebersberger et al. (2000a) and Ebersberger (2002). We will give only a short summary here.

Property 1 *Genetic programming is non-parametric in the sense that it does not require* a priori *assumptions about the structure of the model.*

The researcher using genetic programming does not have to supply an assumption about the structure of the model. Rather, only the components, stored in the primitive function set F and the terminal set T, have to be supplied.

Additionally, only a few parameters concerning the run of the algorithm have to be supplied, such as number of models in the population, mutation probability, cross-over probability and so on. To define the termination of the genetic programming routine the researcher has to provide the maximum computation time (maximum number of iterations) and a satisfactory level of model fit.

Property 2 *Genetic programming only includes relevant components in the models. While improving models, genetic programming disregards irrelevant components.*

The genetic programming routine finds improved models as it constantly evaluates existing models and modifies them. As the models with the most favorable fitness have higher probability of passing their components on to the next population, the components causing a less favorable fitness slowly die out in the whole population.[11]

Property 3 *Genetic programming can be set up to generate models that comply with the prior knowledge about the modeled phenomenon.*

The fitness function in a genetic programming analysis is used to evaluate the models. The quality of a model not only relates to how well the model fits the available observations. If designed appropriately, it also takes into account whether or not the model conforms with the prior knowledge. Hence it is the definition of the fitness function that ensures that the final result complies with the prior knowledge.

Property 4 *Genetic programming generates models that allow for different types of heterogeneity.*[12]

This property follows directly from Property 3. It refers generally to the ability of genetic programming to derive different models for different groups of observations. It also refers to the fact that if a model describes the action of some actors, genetic programming is capable of integrating a component that measures the actors' performance (see, for example, Ebersberger et al. 2000b).

Property 5 *Different applications of genetic programming in scientific research in fields other than economics suggest that genetic programming can handle a variety of types of observations, where the variety of types not only refers to the structure of the data (cross-section, time-series, or panel) but also to the type of the variables (binary, categorical, numerical, logical).*

Property 5 depends on the definition of the primitive function set and the terminal set such that the closure property is maintained.

Property 6 *Genetic programming is capable of detecting structural breaks in the observations and incorporating them in the improved models.*

Property 6 follows directly from a combination of Property 4 and Property 5.

4. IMPROVING A MODEL BY MEANS OF GENETIC PROGRAMMING

In this section we give a brief summary of a modeling exercise that uses genetic programming. A detailed discussion of the exercise can be found in Cantner et al. (2001) and Ebersberger (2002, ch. 10).

4.1 The Initial Model

A recently introduced stylized fact of economic growth refers to the bimodal shape of the distribution of per capita income. Pyka et al. (2003) provide an empirical and theoretical analysis on the evolution of the bimodal distribution. Drawing on the data provided by Summers and Heston (1991), they illustrate the evolution of the twin-peak structure by kernel density estimates.

In order to explain the emergence of a bimodal income structure, Pyka et al. (2003) introduce a theoretical model which shows how differences in the ability of countries to improve their technological performance lead to a clear separation of technologically leading countries and those lagging behind.

In particular, the model consists of three components[13] that drive the development of the bimodal distribution.

1. *Exploitation of intensive technological opportunities.* This is modelled by

$$p(x) = \psi_1\, x\, e^{-\psi_2 x}, \tag{4.1}$$

where p describes the probability of improving from technological level x to the adjacent level by exploiting the intensive opportunity space.

2. *Technological infrastructure:*

$$r(x) = (\psi_3 \cdot x) \cdot (1 + e^{\psi_4 (\psi_5 - x)})^{-1} \tag{4.2}$$

where r is the probability of improving from technological level x to the adjacent level by building on the established technological infrastructure.

3. *Technological obsolescence:*

$$o(x) = \psi_6\, x, \tag{4.3}$$

where $0 < \psi_6 < 1$. o is the probability of the falling behind from x to the adjacent level on the relative technological scale.

Pyka et al. (2003) use the parameters given in Table 4A.1 (in the appendix to this chapter). They combine the three components in a simulation analysis to explain the evolution of the bimodal structure.

4.2 The Need for Improvement

The discussion in Cantner et al. (2001) and particularly in Ebersberger (2002) shows that although the model components yield a bimodal result for world income distribution, they do not fit the available observations in Summers and Heston (1991). Hence each of the model(component)s introduced in Pyka et al. (2003) can be improved by genetic programming.

4.3 Improving the Model by Means of Genetic Programming

Cantner et al. (2001) and Ebersberger (2002) contain a detailed account of the model[14] improvement supported by genetic programming.[15] The language used in this exercise is purely mathematical as $F = \{Plus, Minus, Times, Divide, Power, Log\}$ and $T = \{x, \Re\}$, where $\Re \in [0,1]$ is a random constant.[16]

The improvements of the model components derived with genetic programming are

$$pr(x) = x - x^{\phi_1 + x} \tag{4.4}$$

and

$$o(x) = \phi_2 \cdot x^{\phi_3} \tag{4.5}$$

Econometric analysis in Cantner et al. (2001) and Ebersberger (2002) shows that the obtained model for improvement of the technological position (4.4) and the model for the deterioration of the relative technological position (4.5) yield highly significant parameter estimations.[17]

The model's fit to the data could be slightly improved in the case of *pr*. In the case of obsolescence *o*, the improved model shows an improved fit with the observations.

1. The improvement by means of genetic programming yields model components that show an improved fit to the available observations.
2. The resulting components are more parsimonious than the initial formulations.
3. Due to the parsimony of the components' formulation, interpretation of the results does not cause any difficulties.

5. CONCLUSION

The hypothesis of this chapter is that empirical analysis in economics can profit from the insights and the modeling techniques developed for modeling the learning of bounded rational agents.

In particular we hypothesize that genetic programming is an instrument that supports the process of improving the perception of economic reality. This chapter established a rationale for using genetic programming to improve models in economics. Improvement here refers to an improved fit of the model to real-world observations. Genetic programming is capable of improving models as it can be interpreted as simulating the learning of a bounded rational economist. In doing so, genetic programming offers an empirical approach to data-guided modeling.

The properties sketched here suggest that genetic programming is especially suitable for applications in the context of evolutionary economics. In particular genetic programming

- is non-parametric;
- includes relevant components in the models;
- can include prior knowledge about the problem domain in the models;
- allows for heterogeneity;
- handles various types of observations;
- can detect structural breaks.

To illustrate the use of genetic programming, this chapter summarizes a model improvement exercise carried out by means of genetic programming. Supplementing the theoretical considerations, the result of the model improvement suggests the applicability of genetic programming beyond modeling of bounded rational agents. The economic modeling exercise supported by genetic programming yielded models that, with advantageous properties such as improved fit to the available observations, increased parsimony and interpretability of the models.

APPENDIX

Table 4A.1 Parameters used in Pyka et al. (2003)

Equation	Parameter	Value
(4.1)	ψ_1	0.775
(4.1)	ψ_2	3.500
(4.2)	ψ_3	0.225
(4.2)	ψ_4	15.000
(4.2)	ψ_5	0.500
(4.3)	ψ_6	0.270

Table 4A.2 Parameters for the improved model

Equation	Parameter	Value
(4.4)	ϕ_1	1.545
(4.4)	ϕ_2	0.194
(4.5)	ϕ_3	0.524

NOTES

1. Applications in evolutionary economics are found, for example, in Bhargava and Mukherjee (1994), Wirl (1998) or Keilbach (2000).
2. See, for example, Sommer (1999) or Metcalfe and Calderini (1998).
3. See, for example, Pyka (1999) or Cantner and Pyka (2001).
4. See, for example, Chen and Yeh (1999, 2000), Dosi et al. (1999), or Edmonds (1999a).
5. This general statement does not imply the type of language used for the description of reality. Different types of language, such as mathematical language, natural language etc., may be used for different modeling purposes (Zschocke 1995, p. 97).
6. The online bibliography maintained by W. Langdon and J. Koza represents a nearly complete collection of papers using genetic programming: http: //liinwww.ira.uka.de/bibliography/Ai/genetic.programming.html
7. In most of the models using genetic programming these constructs have a tree-like structure.
8. Here we do not specify where the economist derived those models from.
9. As this is a very general case, domain-specific restrictions or extensions of the language definitions may apply.
10. See, for example, Ebersberger et al. (2000a) or Ebersberger (2002, ch. 8.3).
11. See, for example, a small simulation in Ebersberger (2002, p. 188). Francone (2000, p. 21)

argues that this property can be explicitly used to determine the most important components of the models.

12. For a discussion of different types of heterogeneity see, for example, Cantner and Hanusch (2001) or Cantner (1996).
13. According to the discussion of properties of models in Section 2.1, all three components can be regarded as models.
14. Due to the structure of the observations, component (a) and component (b) had to be combined and are denoted *pr*. Thus the number of models to be improved is reduced to two.
15. The set-up of the genetic programming and the source code can be found in Ebersberger (2002), appendix.
16. To establish closure here the meta-rules supplied by the basic laws of arithmetic and algebra have to be used. Closure in the context of this primitive function and this terminal set means redefining *Divide*, *Power* and *Log* to extend their domain to real numbers (Koza 1992, p. 81).
17. The parameters can be found in Table 4A.2 in the appendix. See also Ebersberger (2002, p. 251).

REFERENCES

Banzhaf, W., Nordin, P., Keller, R.E. and Francone, F.D. (1998), *Genetic Programming – An introduction: On the Automatic Evolution of Computer Programs and Its Applications*, San Francisco, CA: Morgan Kaufmann.

Baumol, W.J. and Blinder, A.S. (1999), *Economics: principles and policy*, 8th edn, San Diego, CA: Harcourt Brace Jovanovich.

Bhargava, S.C. and Mukherjee, A. (1994), 'Evolution and technological growth in a model based on stochastic cellular automata', in L. Leyersdorff and P. v. Besselaar (eds), *Evolutionary Economics and Chaos Theory*, London: Pinter, pp. 55–62.

Blaug, M. (1998), 'Disturbing currents in modern economics', *Challenge* **41**, 11–34.

Boulding, K.E. (1991), 'What is evolutionary economics?', *Journal of Evolutionary Economics*, **1**, 9–17.

Cantner, U. (1996), *Heterogenität und Technologische Spillovers: Grundelemente einer ökonomischen Theorie des technischen Fortschritts*, Habilitationsschrift, Wirtschafts- und Sozialwissenschftliche Fakultät der Universität Augsburg, Augsburg.

Cantner, U., Ebersberger, B., Hanusch, H., Krüger, J.-J. and Pyka, A. (2001), 'Empirically based simulation', *Journal of Artificial Societies and Social Simulation*, **4**. http: //www.soc.surrey.ac.uk/JASSS/4/3/9.html

Cantner, U. and Hanusch, H. (2001), 'Heterogeneity and evolutionary dynamics – empirical conception, findings and unresolved issues', in J. Foster and J. S. Metcalfe (eds), *Frontiers of Evolutionary Economics: Competition, Self-Organization and Innovation Policy*, Cheltenham, UK and Northampton, USA: Edward Elgar.

Cantner, U. and Pyka, A. (2001), 'Technologieevolution: Eine Analyse im Rahmen des Knowledge-Ansatzes der Innovationstheorie', in H.W. Lorenz and B. Meyer (eds), *Studien zur Evolutorischen Ökonomik*, Berlin: Duncker und Humblot, pp. 67–96.

Casti, J.L. (1992), *Reality Rules*, Vol. I: *The Fundamentals*, New York: John Wiley and Sons.

Chen, S.-H. and Yeh, C.-H. (1999), 'Genetic programming in the agent-based modeling of stock markets', in D.A. Belsley and C.F. Baum (eds), *Fifth International Conference: Computing in Economics and Finance*, Boston College, MA, USA, pp. 77–99. Book of Abstracts. http: //fmwww.bc.edu/cef99/papers/ChenYeh.pdf

Chen, S.-H. and Yeh, C.-H. (2000), 'Simulating economic transition processes by genetic programming', *Annals of Operations Research*, **97**, 265–86.

Dosi, G. and Egidi, M. (1991), 'Substantive and procedural rationality, an exploration of economic behavior under uncertainty', *Journal of Evolutionary Economics*, **1**, 145–68.

Dosi, G. et al. (1999), 'Norms as emergent properties of adaptive learning: The case of economic routines', *Journal of Evolutionary Economics*, **9**, 5–26.

Ebersberger, B. (2002), *Genetische Programmierung: Ein Instrument zur empirischen Fundierung ökonomischer Modelle*, Wiesbaden: DUV.

Ebersberger, B., Cantner, U. and Hanusch, H. (2000a), 'Functional search in complex evolutionary environments – examining an empirical tool', in I.P. McCarthy and T. Rakotobe-Joel (eds), *Complexity and Complex Systems in Industry*, Warwick, UK: University of Warwick, pp. 167–79.

Ebersberger, B., Cantner, U. and Hanusch, H. (2000b), *Measuring Efficiency Using a Frontier Search Approach*, Vol. 199, University of Augsburg, Department of Economics Working Paper.

Edmonds, B. (1999a), 'Modeling bounded rationality in agent-based simulations using the evolution of mental models', in T. Brenner (ed.), *Computational Techniques for Modeling Learning in Economics*, *Advances in Computational Economics*, Dordrecht: Kluwer Academic Publishers, pp. 305–32.

Edmonds, B. (1999b), *Syntactic Measures of Complexity*, PhD thesis, University of Manchester, Manchester.

Egidi, M., Marris, R., Simon, H.A. and Viale, R. (eds) (1992), *Economics, Bounded Rationality and the Cognitive Revolution*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Foss, N.J. (1994), 'Realism and evolutionary economics', *Journal of Social and Evolutionary Systems*, **17**, 19–40.

Francone, F. D. (2000), *Discipulus Owner's Manual*, Littleton Register Machines Technologies Inc.

Greene, C.A. (2000), 'I am not, nor have I ever been a member of a data-mining discipline', *Journal of Economic Methodology*, **7**, 217–30.

Hodgson, G. M. (1998), 'Evolutionary economics', in J.B. Davis, D.W. Hands and U. Mäki (eds), *Handbook of Economic Methodology*, Cheltenham, UK and Lyme, USA: Edward Elgar, pp. 160–67.

Keilbach, M. C. (2000), *Spatial knowledge spillovers and the dynamics of agglomeration and regional growth*, Heidelberg: Physika-Verlag.

Koza, J. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Cambridge, MA: MIT Press.

Koza, J.R. (1994), *Genetic Programming II – Automatic Discovery of Reusable Programs*, Cambridge, MA: MIT Press.

Metcalfe, J.S. and Calderini, M. (1998), 'Compound learning, neural nets and the competitive process', *Economics of Innovation and New Technology*, **7**, 271–302.

Moss, S. and Edmonds, B. (1998), 'Modeling learning as modeling', *Cybernetics and Systems*, **29**, 5–37.

Pyka, A. (1999), *Der kollektive Innovationprozess: Eine theoretische Analyse informeller Netzwerke und absorptiver Fähigkeiten*, Berlin: Duncker und Humblot.

Pyka, A., Krüger, J.J. and Cantner, U. (2003), 'Twin peaks – what the knowledge-based approach can say about the dynamics of the world income distribution', in P.P. Saviotti (ed.), *Applied Evolutionary Economics: New Empirical Methods and Simulation Techniques*, Cheltenham, UK and Northampton, USA: Edward Elgar, ch. 9.

Sommer, J. (1999), *Künstliche neuronale Netze zur Beschreibung des Anlegerverhaltens auf spekulativen Märkten: Eine theoretisch-experimentelle Analyse*, Berlin: Tenea.

Summers, R. and Heston, A. (1991), 'The Penn World Table (mark 5): An expanded set of international comparisons 1950–1988', *Quarterly Journal of Economics*, **106**, 327–68.

Williams, M. (1999), 'The philosophy of economic modeling: a critical survey', *South African Journal of Philosophy*, **18**, quoted here from: mimeo, Department of Economics and Social Sciences, De Montfort University. Leicester. Stand 8.12.98. http://www.mk.dmu.ac.uk/~mwilliam.

Wirl, F. (1998), 'Socio-economic typologies of bureaucratic corruption and implications', *Journal of Evolutionary Economics*, **8**, 199–220.

Zschocke, D. (1995), *Modellbildung in der Ökonomie*, München: Vahlen Verlag.

Zwicker, E. (1981), *Simulation und Analyse dynamischer Systeme in den Wirtschafts- und Sozalwissenschaften*, Berlin and New York: de Gruyter.

5. Entropy statistics as a framework to analyse technological evolution

Koen Frenken and Alessandro Nuvolari*

1. INTRODUCTION

Many scholars have suggested that important similarities exist between technological development and biological evolution and that, for this reason, evolutionary models can provide us with fairly adequate representations of technical change (Nelson and Winter 1982, Basalla 1988, Mokyr 1990). However, as has been repeatedly pointed out by those who endorse the adoption of an evolutionary approach, there are also substantive differences between biological evolution and technological evolution (Freeman 1991, Nelson 1995). Therefore, evolutionary models should always be employed with caution, taking into account the specificities of the processes of mutation and selection under study.

The issue we are considering here concerns evolutionary processes of a special kind, namely the way complex entities evolve through processes of mutation and selection. Recent evolutionary theorizing in biology and artificial intelligence has stressed that complex entities evolve in ways that are different from non-complex ones in important respects. This claim also has significant implications for models of technological evolution, as a technological artefact is a complex evolving entity *par excellence* (Rosenberg 1976).

Following Simon's (1969 [1996]) work on the design of artificial systems, we describe a technological artefact as a man-made system constituted by interconnected components that are intended to collectively perform a number of functions. The complexity of an artefact is due to the interdependencies between components, which causes only some combinations of elements to work well together, in the sense that these combinations are capable of achieving satisfactory levels of performance. In Simon's view, a good deal of what we call innovative activities consists of trying to improve the general performance of the artefact by finding out progressively better configurations of its constituting elements.[1]

Until recently, however, formal treatments of system interdependencies for the understanding of technological innovation have been scarce. This has changed with the introduction of 'complexity' models from natural sciences in the realm of (evolutionary) economics. In this respect, Kauffman's (1993) NK model of evolutionary biology has proven extremely promising and has already been adopted in a large number of contributions in the innovation and organization literature.[2] The NK model represents the design process of a complex technological artefact as a trial-and-error process that is bound to end up in a local optimum. Although the NK model has received considerable attention, much less effort has so far been put into empirical applications.[3] In this chapter, we set out a framework based on entropy statistics, which allows a relatively straightforward application of the NK model to empirical studies of technological change.

We consider the examples of the early development of the steam engine (1760–1800), the development of the aircraft (1913–84), and the development of the helicopter (1940–83) to illustrate the way in which the NK model can be employed in empirical studies of technical change by means of entropy statistics. As we will see, the interpretative accounts that we were able to produce using the NK model in combination with the entropy methodology emend the received histories of the technologies we are examining in this chapter in important respects. This suggests that other historical studies of technology could indeed benefit greatly from the adoption of the type of approach we propose in this study.

The remainder of the chapter is organized as follows. Section 2 contains an exposition of Kauffman's (1993) NK model and a number of generalizations since developed. Section 3 presents our entropy methodology in detail. Section 4 applies the entropy framework to data on steam engines, aircraft, and helicopters, and discusses the results in the light of received histories of these three technologies. Section 5 draws conclusions.

2. TECHNOLOGICAL DEVELOPMENT AS A SEARCH PROCESS ON RUGGED LANDSCAPES

Many scholars have recognized that interdependencies between components in technological artefacts are the prime source of design complexity (Simon 1969 [1996], Rosenberg 1976, Sahal 1985, Vincenti 1990, Ziman 2000). The existence of interdependencies between components implies that the functioning of a system cannot be fully understood from the functioning of its individual components. Depending on the precise combination of the components that make up a system, a component will function in a different way. And, each time one manages to improve the functioning of one component, new problems can arise in other components accordingly requiring redesign. In this context, Rosenberg (1976) introduced the concept of 'technical imbalances' between components that trigger sequences of problems and solutions over time.[4]

The existence of system interdependencies is what we understand to be the nature of complexity in the development of new technological designs. In this perspective, the design task essentially consists of the combinatorial problem of assembling the right set of components in a functioning system. The space of all the possible combinations between all the possible configurations of all the components of a system is called the 'design space' of a technology (Bradshaw 1992). Assume that a technology can be described by N components, or more generally, dimensions ($i=1,\ldots,N$). Along each dimension i there exist A_i possible states or configurations, called 'alleles', which can be coded as '0', '1', and so on. Each possible design can then be written as a string of alleles $s_1 s_2 \ldots s_N$ and is part of N-dimensional design space S, for which it holds that:[5]

$$s \in S;\ s = s_1 s_2 \ldots s_N;\ s_i \in \{0, 1, \ldots, A_i - 1\}. \qquad (5.1)^6$$

The combinatorial nature of a design space implies that the size of a design space increases exponentially for linear increases in N. The size of the design space S is given by the product of the number of alleles along each dimension:

$$S = \prod_{i=1}^{N} A_i \qquad (5.2)$$

In the case of binary strings (that is, when all dimensions contain only two alleles '0' and '1'), the size of design space equals $S = 2^N$, meaning that the number of possible designs doubles for each dimension added. As technological artefacts are typically made up of many dimensions and many alleles per dimension, they have enormous design spaces. Exploring the whole design space would obviously be very expensive. Instead designers will usually apply search rules that allow them to economize by examining only subsets of the design space. Thus only a small part of the design space will in effect be searched, and an even smaller part of the design space will be commercialized on product markets.

2.1 The NK Model

Kauffman and Levin (1987) and Kauffman (1993) developed the NK model to examine the properties of evolving complex systems with varying degrees of complexity. Complexity stems from interdependencies between

	$i = 1$	$i = 2$	$i = 3$
w_1	**x**	x	–
w_2	x	**x**	–
w_3	x	–	**x**

Figure 5.1 Example of an architecture of an N *=3 system with* K *=1*

the constituting dimensions of a system, such as genes in biological organisms and components in technological artefacts. The interdependencies between dimensions in a complex system are called 'epistatic relations'. An epistatic relation between components implies that when a component mutates, the mutation affects not only the functioning of the component itself but also the functioning of all the components that are 'epistatically related' to it. The ensemble of epistatic relations in a technological system is called a technology's architecture (Henderson and Clark 1990).[7]

The NK model is restricted to particular types of system architectures that can be expressed by a single parameter K, which stands for the number of other components that affect the functioning of each component. For example, the class of systems for which $K = 1$ holds refers to systems with an architecture in which the functionality of each component depends on the choice of allele of the component itself and on the choice of the allele of one other component. The K parameter can be considered an indicator of a system's complexity, with $K = 0$ being the least complex and $K = N - 1$ the most complex architecture. When $K = 0$ each technical dimension is independent of any other dimensions. Optimization can then proceed by optimizing each individual dimension separately, which will lead automatically to the global optimum. For increasing values of K it will become increasingly hard to optimize the system design globally, as interdependencies exist between dimensions. The number of local optima in which one can end up increases with the value of K.[8]

Consider, as an explanatory example, a system for which $N = 3$ and $K = 1$ hold, with an architecture as specified in Figure 5.1. Mutations in components in the columns affect the functioning of the component in the row as indicated by 'x'. The symbol '–' denotes that there is no epistatic relation between the component in the row and the component in the column. The architecture in Figure 5.1 specifies the following epistatic relations between the three components in the system. The functioning of the first component w_1 changes when the first component itself or the second component is mutated. The functioning of the second component w_2 changes when the

	w_1	w_2	w_3	W
000:	0.7	0.2	0.3	0.40
001:	0.7	0.2	0.6	0.50
010:	0.9	0.3	0.3	0.50
011:	0.9	0.3	0.6	0.60
100:	0.8	0.7	0.1	0.53
101:	0.8	0.7	0.4	0.63
110:	0.3	0.5	0.1	0.30
111:	0.3	0.5	0.4	0.40

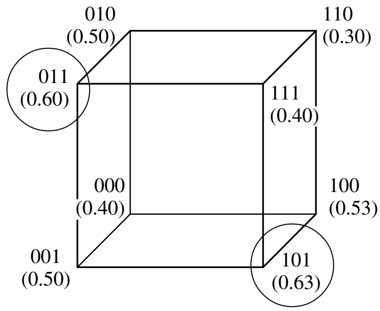


Figure 5.2 Simulation of a fitness landscape of an N *=3 system with* K *=1*

second component itself or the first component is mutated. And the functioning of the third component w_3 changes when the third component itself or the first component is mutated.

Following Kauffman (1993), we construct a fitness landscape by drawing randomly the value of the fitness w_i of component i from the uniform distribution between 0 and 1. A random value is drawn for w_i each time component i is itself mutated and each time another component that epistatically affects component i is mutated. System fitness W is derived as the mean value of the fitness values of all components:

$$W(s) = \frac{1}{N} \cdot \sum_{i=1}^{N} w_i\,(s_i) \tag{5.3}$$

A simulation of a fitness landscape is given in Figure 5.2. The circled strings are local optima or 'peaks' on a 'rugged fitness landscape'. For these local optima it holds that all neighbouring strings, that is, the strings that can be reached by a mutation in one component, have a lower fitness W. In the simulation in Figure 5.2, this property holds for strings 011 and 101 as their system fitness values W(011) and W(101) exceed the values of their neighbouring strings. Local optima reflect complementary alleles as the collective fitness exceeds the value of neighbouring strings.

Using the concepts of design space and fitness landscape, the design process can be modelled as a local search process based on trial and error. Local search proceeds by means of a mutation in one, randomly chosen, dimension (a *trial*). A mutation means that a designer moves to a neighbouring string in the design space. The newly found string is accepted when system fitness W increases, while it is rejected when system fitness decreases (*error*). Acceptance of a mutation implies that search continues from the newly found string, and rejection implies that search continues from the previous string. In this way, a designer can search for improvements in an incremental way until a local optimum is found that can no longer be improved by means of a mutation in one dimension.[9] Trial-and-error search can thus be considered as an 'adaptive walk' over a fitness landscape towards a local optimum, and the search will only halt when a local optimum is reached. Following the metaphor of the fitness landscape, this type of search in complex technological systems can be considered a process of 'hill-climbing'.

It should be stressed that we used the relatively simple case of $N=3$ in the example above for explanatory purposes. In real-world R&D activities, the number of design dimensions N is generally much larger. Consequently, local search takes place in much larger design spaces containing many more local optima for the same value of K. The probability of ending up in a local optimum is correspondingly much higher.

An important property of the NK model holds that the number of local optima in a fitness landscape is a function of the complexity K of a system's architecture. When complexity is absent ($K=0$), the fitness values of each dimension are not affected by mutations in other dimensions. Therefore, the global optimum of a system of $K=0$ can always be found by local search through trial and error as described above. Put another way, fitness landscapes of $K=0$ systems always contain only one optimum (which is by definition the global optimum). For systems with a positive K value, the fitness values of dimensions are affected by mutations in dimensions that are epistatically related. As a result, the fitness landscape will generally contain multiple local optima. Kauffman (1993) has shown that the expected number of local optima increases for increases in K. This means that it becomes increasingly hard to find the global optimum for systems with higher complexity.

A second property of the NK model holds that the fitness of local optima decreases for increases in K. One can understand this outcome as reflecting the detrimental effects of a higher number of conflicting constraints between components. The higher K, the more difficult it becomes to improve the fitness of one component without lowering the fitness of other components. Consequently, the system fitness of local optima is generally quite low. Furthermore, the variance of fitness value of local optima also decreases for increases in K, which means that the differences in fitness of local optima become smaller for systems with higher complexity. In the context of competing technological designs, this result suggests that the higher a technology's complexity, and the smaller the performance differences between locally optimal designs, the more persistent design variety will be.[10]

2.2 Generalizations of the NK Model

The NK model can be generalized in a number of respects to represent a wider range of phenomena. The first generalization concerns the representation of the relation between a system's 'genotype' (the set of design dimensions) in relation to the 'phenotype' (the set of functions a system performs). The NK model is based on the idea that each component of the system performs an 'own' sub-function within the system with regard to the attainment of one overall function on which external selection operates (Kauffman 1993, p. 37). Each component i is conceived to have a particular fitness value w_i that reflects its functional contribution to the system as a whole. The fitness of the system as a whole is derived as the average of the fitness of individual components.

Altenberg (1994, 1995, 1997) describes a generalized (biological) model of complex systems that contains N dimensions ($i=1,\ldots,N$) and F functions ($f=1,\ldots,F$) and for which it holds that N does not necessarily equal F. In biological systems, for which the original and generalized NK models were both initially conceived, an organism's N genes are the system's components and an organism's F traits are the system's functions on which natural selection operates. The string of genes constitutes an organism's genotype and the set of traits constitutes an organism's phenotype. The genotype of an organism is the level at which mutations take place, which are transmitted to offspring. The phenotype is the level at which natural selection operates in terms of its relative fitness.

Analogously, a technological artefact can be described in terms of its N components and the F functions it performs. The string of alleles describes the 'genotype' of a technological system, and the list of functions describes the 'phenotype' of this system. Typical functions of technological artefacts include cost-related criteria (fuel-efficiency, maintenance cost, and so on) and performance-related criteria (power, speed, weight, safety and the like).[11]

In Altenberg's generalized NK model, the architecture of a complex system is represented by a 'genotype–phenotype matrix' of size $F \times N$ with:

$$M=[m_{fi}], f=1, \ldots, F, i=1, \ldots, N \tag{5.4}$$

As in the NK model, an epistatic relation is represented by 'x' when function f is affected by component n and by '–' when function f is not affected by the component n. An example of a matrix for $N=3$ and $F=2$ is given in Figure 5.3.

The way in which fitness landscapes are constructed for generalized genotype–phenotype matrices follows the same logic as the original NK model

	$i = 1$	$i = 2$	$i = 3$
w_1	x	x	–
w_2	–	x	x

Figure 5.3 Example of a generalized genotype–phenotype matrix

discussed in the previous section. For each component that is mutated, all functions that are affected by this component are assigned a new, randomly drawn fitness value w_f from the uniform distribution between 0.0 and 1.0. Total fitness W is again derived as the mean of the fitness values of all functions:

$$W(s) = \frac{1}{F} \cdot \sum_{f=1}^{F} w_f(s) \tag{5.5}$$

A simulation of the fitness landscape example of the genotype–phenotype matrix of Figure 5.3 is given in Figure 5.4 for all possible combinations between two alleles of three components. Local optima are again circled, reflecting the combinations in which component technologies are complementary.

The meaning of the concepts of fitness landscape and local optima remains entirely the same in Altenberg's (1994, 1995, 1997) generalized NK model. Moreover, the properties of the NK model discussed above, which relate the number of local optima, the fitness of local optima, and the variance in fitness of local optima to complexity K, remain intact. The main difference compared to the original NK model is that in the generalized model the number of dimensions N is not necessarily equal to the number of functions F. Altenberg's (1997) model can therefore be considered as an important generalization of the original NK model of complex systems by Kauffman (1993).[12]

A second generalization of the NK model can be introduced by specifying a more general fitness function that translates the fitness levels of individual functions w_f into one overall assessment value W. The fitness function in Equation (5.5) specified that each function is weighted equally. As an empirical specification of fitness (performance) of a technology, this equation obviously does not account for the general case in which users may apply different weights to the various functions of the artefact. Allowing for different values of weights for each function, we get:

$$W(s) = \sum_{f=1}^{F} \beta_f \cdot w_f(s) \tag{5.6}$$

$$\sum_{f=1}^{F} \beta_f = 1, \beta_f > 0 \tag{5.7}$$[13]

A selection environment can then be defined by the set of weights $\{\beta_1, \beta_2, \ldots, \beta_F\}$ that is applied by users of the technology. The concept of a fitness landscape does not change when total fitness is computed as a weighted sum instead of as the average of the fitness values of functions. However, the values of total fitness of each design $W(s)$ will be different depending on the values of the weights that are applied.

A final generalization can be introduced by allowing for heterogeneity among users. So far, we have implicitly assumed that each user of a particular design applies the same set weights and thus assigns the same fitness value $W(s)$ to a design. However, depending on the specific use of the design, different users may well apply different weights, and thus assign different fitness values to one and the same design (Lancaster 1966, 1979, Saviotti and Metcalfe 1984, Saviotti 1996).[14] In this case of heterogeneous demand, different users have different valuations of the same technological design, as they weight the levels of functions differently. As Lancaster (1979, p. 17) expressed it:

> Differences in individual reactions to the same good are seen as expressing different preferences with respect to the collection of characteristics possessed by that good and not different perceptions as to the properties of the good.

The weights assigned to functions as specified above $\{\beta_1, \beta_2, \ldots, \beta_F\}$ reflect one homogeneous user group. When there is more than one user group, we

	w_1	w_2	W
000:	0.8	0.9	0.85
001:	0.8	0.6	0.70
010:	0.4	0.3	0.35
011:	0.4	0.2	0.30
100:	0.2	0.9	0.55
101:	0.2	0.6	0.40
110:	0.9	0.3	0.60
111:	0.9	0.2	0.55

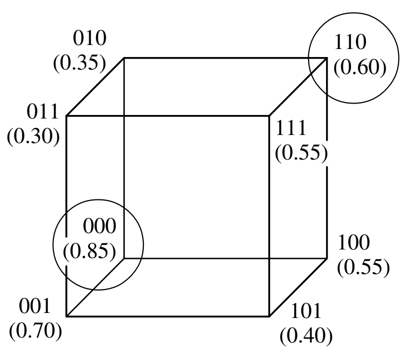


Figure 5.4 Simulation of fitness landscape of the matrix in Figure 5.3

can characterize each different user group by a different set of weights. For a G number of user groups g ($g = 1,\ldots,G$), we have G sets of weights. For each user group, the fitness W_g of a design is given by:

$$W_g(s) = \sum_{f=1}^{F} \beta_{fg} \cdot w_f(s) \tag{5.8}$$

$$\sum_{f=1}^{F} \beta_{fg} = 1,\ \beta_{fg} \geq 0$$ (5.9)[15]

This specification of the selection environment includes the specification given above for a homogeneous selection environment as the special case in which $G = 1$.

When heterogeneity in preferences is more dispersed, it is less likely that one design is optimal for all user groups. In that case, product differentiation is expected to occur. In the extreme case, given a sufficiently large design space, a different design may be found for each different user group. When a user group exists for which no design is yet optimized, and this is known to designers, this in itself can spur the search for innovations in particular components in order to find a new design capable of fitting their particular demand (induced innovation).

2.3. IMPLICATIONS

Regarding the patterns of technological evolution one can expect to emerge, a number of implications follow from the previous discussion of the NK model and its generalizations. To summarize the implications, one can distinguish between technologies that are subject to homogeneous demand and technologies that are subject to heterogeneous demand:

- When demand is homogeneous ($G = 1$) and complexity is absent ($K = 0$) there exists only one local optimum, which can be found by local trial and error. When demand is homogeneous ($G = 1$) and system complexity is present ($K > 0$), the expected number of local optima becomes a function of the complexity parameter K. Thus, even when demand is homogeneous, technological variety is expected to emerge as different designers will come up with different, locally optimal solutions. What is more, the variety in technological designs can be quite persistent for highly complex architectures, as the variance in the fitness values of local optima has been shown to decrease as a function of K (Kauffman 1993).
- When demand is heterogeneous ($G > 1$) and complexity is absent

($K = 0$), there is only one global optimum, which is the same for each user group, since each function can be optimized independently from other functions. The existence of heterogeneous demand is not a sufficient condition for design variety to emerge. When demand is heterogeneous ($G > 1$) and complexity is present ($K > 0$), design variety is expected to emerge for two reasons. First, as in the case of homogeneous demand, trial and error may lead different designers to come up with different local optima. Second, the heterogeneity in preferences may render different designs to be globally optimal for different user groups. Note that in the case of heterogeneity in preferences, the design variety is expected to be even more persistent than in the case of homogeneity in preferences where variety may slowly disappear as sub-optimal design lose ground due to small differences in fitness.

Note that design variety that is expected to emerge following the generalised NK-model is always limited by the extent to which scale economies and network externalities are realized in the production and use of a single design (Arthur 1989). When one design s is produced and used in much higher numbers than alternative designs, lower price and higher willingness -to-pay may attract users who previously preferred an alternative design. However, a greater degree of heterogeneity of preferences as expressed by the different weights users assign to the various functions, will in turn render it less likely that one design will attract all users. Moreover, radical innovation may at all times lead to the introduction of complete new designs attracting new users or users who previously adopted another design.

Patterns of technological evolution thus depend crucially on the complexity of a technology's architecture, the number of functions that can be distinguished, and the degree of heterogeneity of demand. In the three cases we will discuss below (steam engines, aircraft, and helicopters), the complexity and number of selection criteria is generally estimated to be quite high (Rosenberg 1982). Moreover, all three technologies have been used in a wide range of user contexts. In our view, one can expect an empirical analysis to show technological variety to emerge in the course of their evolution.

3. ENTROPY STATISTICS

In the previous section we proposed a formalization of artefact complexity and discussed its implications for the patterns of technological evolution that are expected to emerge. A straightforward way to analyse empirical data on artefact designs in terms of the ruggedness-of-fitness landscapes is to apply entropy statistics. Entropy statistics can be computed using frequency distributions of technological designs coded in the N-dimensional design space and they allow one to map both the degree of technological variety (by means of entropy indices), and the nature of technological variety (by means of mutual information indices). In this way, evolutionary trends in the development of a technology can be consistently outlined.

The entropy index refers to the degree of randomness in the choice of technological designs as reflected by the skewness of a distribution. A skewed distribution reflects a situation in which designers hardly differ in their choice of design, while a flat distribution reflects a situation in which designers have come up with very many different designs. As such, entropy can be used as an indicator of technological standardization and to what extent a dominant design can be said to have emerged (Frenken et al. 1999b). The more skewed a distribution, the lower the entropy (randomness) of a distribution.

To understand to what extent the variety indicated by entropy can indeed be said to reflect local optima on a rugged-fitness landscape, a second indicator called mutual information is introduced (Frenken 2000, 2001). Mutual information indicates the extent to which particular alleles along different dimensions co-occur in the technological designs offered on the market. Statistically, mutual information thus indicates the degree of dependence between different design dimensions. The existence of local optima would imply that particular alleles along one dimension typically co-occur often with particular alleles along other dimensions, which would result in a high value of mutual information (dependence). Following the metaphor of a fitness landscape, high mutual information indicates that designers occupy more than one peak. When alleles along different dimensions are more or less randomly combined, mutual information is low (independence). Designers are more or less randomly spread out over the fitness landscape without clustering around specific peaks.

3.1 Entropy

The entropy concept was developed in late nineteenth-century thermodynamics to describe randomly moving particles (Prigogine and Stengers 1984). When many particles are moving randomly through a state space, like particles of a gas in a box, the resulting distribution of all particles is completely flat. The flat distribution follows from the fact that at all times each particle has an equal probability of being present in any area in the box. The flat distribution is characterized by maximum entropy (randomness). When particles behave in a non-random way, some areas in the box will be filled with more particles than other areas, and the resulting distribution is skewed. In that case, the entropy of the distribution is lower compared to the case in which all particles move randomly. In the extreme case when all particles cluster in one area of the box, entropy is lowest.

Entropy is thus a macroscopic measure at the level of a distribution that indicates the degree of randomness in the microdynamics underlying a frequency distribution. As such, entropy can also be used as a variety measure of frequency distributions of technological designs. Following Saviotti (1996), we refer to a distribution of technological designs as the 'product population'. Maximum entropy corresponds to the case in which all designs occur at the same frequency. Such a completely flat distribution would occur when designers move around randomly in state space, which has been called here the 'design space'. In that case, designers pick randomly the various alleles of each component. In this hypothetical case, any product design has an equal probability of occurrence, and the product population would be characterized by even frequencies of all designs. This hypothetical situation refers to a situation in which designers do not learn about the functional properties of different designs, and simply choose the alleles configuration at random (analogous to the randomly moving particles in a box, explained above). A skewed distribution occurs when some designs dominate the product population. In that case, the frequency of some designs is high, while the frequency of most designs is low or zero. In this case, designers have not chosen a design at random, but have somehow learned which designs are most demanded, for example, by applying a local search strategy of hill-climbing. In the extreme case in which all designers choose to offer one and the same design on the market, entropy will be minimum.

The entropy measure thus indicates the degree of design variety in a product population. To describe a product population as a frequency distribution of designs, let each design be coded again as a string of N alleles ($i = 1,\ldots,N$). Each of the N dimensions is labelled here as X_i, with each dimension containing A_i alleles again coded as '0', '1', and so on. The relative frequency of design s in the product population is denoted as p_s. The entropy value of an N-dimensional distribution is then given by (Theil 1967, 1972, Langton 1990):

$$H(X_1, \ldots, X_N) = -\sum_{s_1=0}^{A_1-1} \cdots \sum_{s_N=0}^{A_N-1} p_s \cdot \ln p_s. \qquad (5.10)$$ [16,17]

Entropy is zero when all products present in the population are designed according to one and the same design. This design would have a frequency of one in the product population, which implies that the entropy of the product population equals:

$$H_{\min} = -1 \cdot \ln(1) = 0.$$

Entropy is positive otherwise. The larger the entropy value, the larger the design variety in the product population. The maximum entropy is limited by the size of design space S. When all S possible combinations of alleles have an equal frequency, we obtain a uniform distribution in which each design has frequency $p_s = 1/S$. The entropy of this distribution equals:

$$H_{\max} = -S \cdot \left(\left(\frac{1}{S} \right) \cdot \ln \left(\frac{1}{S} \right) \right) = -\ln \left(\frac{1}{S} \right) = \ln(S).$$

This value is the maximum possible entropy value for a distribution of product designs with a design space of S possible designs. It implies that for larger values of S, maximum entropy increases, with the marginal increase of maximum entropy decreasing. This property reflects that each new entity added contributes to variety, but decreasingly so.

Similarly, the design variety along one dimension i can be computed. The one-dimensional or marginal entropy indicates the variety in a product population with respect to one design dimension only, and is given for each dimension by:

$$H(X_i) = -\sum_{s_i=0}^{A_i-1} p_{s_i} \cdot \ln p_{s_i}. \tag{5.11}$$

As we will see, the one-dimensional entropy formula can be used to compute the mutual information index, which is equal to the difference between the sum of one-dimensional entropy values and the N-dimensional entropy value.

3.2 Mutual Information

In information theory, the measure that indicates the degree of dependence (co-occurrence of alleles) in a frequency distribution is the measure of mutual information T. Mutual information is given by (Theil 1967, 1972, Langton, 1990):

$$T(X_1, \ldots, X_N) = \sum_{s_1=0}^{A_1-1} \ldots \sum_{s_N=0}^{A_N-1} p_s \cdot \ln \frac{p_s}{\prod_{i=1}^{N} p_{s_i}}. \tag{5.12}$$

The mutual information value T indicates the extent in which alleles along different dimensions are co-occurring in the distribution of designs. The mutual information value equals zero when there is no dependence between any of the dimensions. In that case, the joint frequency of alleles of components p_s corresponds exactly to the frequency that could be expected from the product of the marginal frequencies $\Pi_{i=1}^{N} p_{s_i}$. When the product of marginal frequencies does not correspond to the joint frequency, there is dependence between dimensions. Mutual information is thus derived by the weighted sum of dependence values for each design. It can be proven that the weighted sum of dependence values is non-negative for any frequency distribution; that is, $T \geq 0$ (Theil 1972). The greater the difference between the joint frequency and the product of marginal frequencies, the higher the value of the mutual information, and the more alleles along particular dimensions co-occur in 'design families'.

The mutual information measure is directly related to the concept of entropy as mutual information can be derived from the multi-dimensional and marginal entropy values. In the general case of an N-dimensional distribution ($N > 1$) the mutual information equals the sum of marginal entropy values minus the N-dimensional entropy value (Theil and Fiebig 1984, p. 12):

$$T(X_1, \ldots, X_N) = \left(\sum_{i=1}^{N} H(X_i) \right) - H(X_1, \ldots, X_N). \tag{5.13}$$

From this equation, it can be derived that the mutual information equals zero if entropy equals zero, and that mutual information equals zero if entropy is maximum (see Appendix).

Similarly, one can compute the mutual information between each pair of dimensions to indicate dependence between two dimensions:

$$T(X_i, X_j) = H(X_i) + H(X_j) - H(X_i, X_j) \quad i \neq j; i = 1, \ldots, N; j = 1, \ldots, N.$$

The two-dimensional mutual information values indicate the dependence between a pair of dimensions and are thus informative with regard to the importance of epistatic relations among the pair of dimension in question. A high mutual information between two dimensions suggests that an important epistatic relation exists between the two dimensions, since designers predominantly offer alleles in particular opposite combinations (for example, either combination 00 or combination 11). Dependence reflects dominant complementarities between two dimensions as particular alleles along the one dimensions often co-occur with particular alleles along the other dimension and irrespective of alleles in yet other dimensions.

3.3 Entropy and Mutual Information as Indicators of Evolution

To explain the connection between entropy and mutual information indicators and the exploration of rugged-fitness landscapes, one should keep in

Table 5.1 Three examples of distribution for binary strings of N = *3*

Distribution	p_{000}	p_{001}	p_{010}	p_{011}	p_{100}	p_{101}	p_{110}	p_{111}
Case 1	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
Case 2	0.250	0.000	0.000	0.250	0.000	0.250	0.250	0.000
Case 3	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.500

Entropy	$H(X_1,X_2,X_3)$	$H(X_1, X_2)$	$H(X_1, X_3)$	$H(X_2, X_3)$	$H(X_1)$	$H(X_2)$	$H(X_3)$
Case 1	ln 8	ln 4	ln 4	ln 4	ln 2	ln 2	ln 2
Case 2	ln 4	ln 4	ln 4	ln 4	ln 2	ln 2	ln 2
Case 3	ln 2	ln 2	ln 2	ln 2	ln 2	ln 2	ln 2

Mutual information	$T(X_1, X_2, X_3)$	$T(X_1, X_2)$	$T(X_1, X_3)$	$T(X_2, X_3)$
Case 1	0	0	0	0
Case 2	ln 2	0	0	0
Case 3	ln 4	ln 2	ln 2	ln 2

mind the relationship between entropy and mutual information. Recall Formula (5.13), which expresses mutual information as the sum of marginal entropy values minus multi-dimensional entropy, which can be rewritten as:

$$\left(\sum_{i=1}^{N} H(X_i)\right) = T\,(X_1, \ldots, X_N) + H\,(X_1, \ldots, X_N)$$

From this formula, it can readily be seen that, given a value for the sum of marginal entropy (ΣH_i), mutual information can increase only at the expense of (total) entropy, and vice versa. This relationship is illustrated in Table 5.1, in which three different frequency distributions of designs are listed (for $N = 3$). In all three cases, the sum of marginal entropy values is the same ($\Sigma H_i = 3 \cdot \ln(2) = \ln(8)$), because in all three cases the two alleles along each dimension occur at the same frequencies. However, the three-dimensional entropy and mutual information values differ for each distribution. Case 1 corresponds to a uniform distribution with maximum entropy and zero mutual information. Case 2 shows that a multi-modal distribution with four designs has positive and equal frequencies. Three-dimensional entropy equals ln(4), while three-dimensional mutual information is only ln(2). Finally, case 3 shows a bi-modal, 50–50 distribution in which two opposite designs are present in the product population (000 and 111). In this case, three-dimensional entropy equals only ln(2), while three-dimensional mutual information adds up to ln(4). The latter case is characterized by such high mutual information because knowledge of one allele along one dimension of a design would allow one to perfectly predict the alleles along the two other dimensions.

When entropy and mutual information are applied to the frequency distributions of consecutive years of technological evolution, a very different picture may emerge. In that case, the value of ΣH_i in a particular year will differ from the value of ΣH_i in other years. Over time, the value of ΣH_i may increase or decrease, or show no trends. An increasing trend would indicate a growing variety in alleles used along each design dimension. Following the formula, such an increase in the value ΣH_i implies that entropy and mutual information can both increase at the same time. In that case, we have a pattern of increasing design variety as indicated by the rise in $H(X_1 \ldots X_N)$ *and* of increasing differentiation of designs in families as indicated by the rise in $T(X_1 \ldots X_N)$. Such a process indicates the progressive development of a growing number of design families akin to 'speciation' in biology (Saviotti 1996, Levinthal 1998). The reverse pattern can also take place. When ΣH_n is falling, entropy and mutual information may decrease at the same time (for example when a product family totally disappears).

The evolutionary development of a complex technology, following the generalized NK model as discussed earlier, is expected to be characterized by both an increasing degree of variety (entropy) and an increasing degree of differentiation (mutual information). Such a development process can be understood from the multi-dimensional and complex nature of technological artefacts and the existence of heterogeneous demand.

4. APPLICATIONS

We will test our thesis of growing design variety and differentiation into design families using data on early steam engines (1760–1800), aircraft (1913–84) and helicopters (1940–83). For each technology we will first provide a short summary of the 'standard' historical account of its development, then present the data and results, and finally discuss what new insights can be derived from the analysis.

4.1 Steam Engines

4.1.1 Early steam-engine history[18]

Historians of technology have described the early development of steam-power technology as a 'linear' succession of technological breakthroughs. The main contours of what might be called the traditional account[19] of early steam-engine development concern the design sequence of Savery–Newcomen–Watt–Trevithick that took place during the eighteenth century.

In the late seventeenth century mining activities begun to be severely hampered by flooding problems. Following the scientific investigations of Torricelli and Pascal, there were several attempts to use atmospheric pressure to lift water out of mines. The Savery engine can be considered as the first successful effort in this direction. The engine was developed during the period 1695–1702. In the Savery engine, steam was first admitted and then condensed inside a 'receiving' vessel by pouring cold water over its outside. Following steam condensation, atmospheric pressure drove water up into the vessel. The engine had two major shortcomings, which limited its practical utilization: restricted height of operation and high fuel consumption due to the need for recreating steam inside the vessel at each stroke.

The Newcomen engine, developed in 1712, resolved the problem of limited height of operation. The Newcomen engine consisted of a piston–cylinder arrangement connected with one hand of a working beam. Steam was admitted from the boiler into the cylinder by means of a valve. Then a cold of jet of water was sprayed into the cylinder, condensing the steam. At this point, because of the creation of a partial vacuum, atmospheric pressure pushed the piston down, lifting the pump rod at the other end of the beam. The use of the piston–cylinder arrangement together with the beam made it possible to use the engine for effective mine drainage. Furthermore, the Newcomen engine was robust, highly reliable and based on a fairly simple working principle. The Newcomen engine, however, did not solve the problem of high fuel consumption. Neither did the engine design deliver smooth motion, preventing the use of this kind of engine in applications in which a smooth rotary motion was needed.[20]

James Watt in the 1770s and in the 1780s successfully tackled these two problems. In his engine, condensation was carried out in a separate vessel and not in the cylinder. This design implied that there was no longer the need to reheat the cylinder at each stroke, which greatly contributed to fuel-efficiency. After the invention of the separate condenser, Watt conceived a number of modifications to his engine in order to allow the effective transformation of reciprocating motion into rotary motion. Among the designs that were developed for rotary motion was the double-acting Watt engine, in which steam is admitted into the cylinder on both sides of the piston in an alternating manner. This resulted in a more powerful action, but also in a much more regular movement of the piston.

Finally, in the second half of the 1790s, Richard Trevithick developed the first high-pressure engine (Watt engines used steam at a little more than atmospheric pressure). This type of engine did not use the separate condenser, but discharged exhaust steam directly into the atmosphere. For this reason, they were called ‘puffers’. The main advantage of this type of engine was their compactness and their cheaper cost of installation due to elimination of the condenser, the air pump and the beam.[21]

As is apparent from this narrative, such a historical depiction is akin to chronicling a sort of ‘glorious march of invention’, where most of the emphasis is put on the creative contributions of a succession of individual inventors (the line Savery–Newcomen–Watt–Trevithick). Each inventor tackled the shortcomings of the technological ‘state of the art’, devising improvements that made previous engine designs obsolete through a process of technological substitution. The question is whether this traditional picture also emerges from entropy analysis.

4.1.2 Early steam-engine data

The data we use are taken from an up-to-date version of the database collected by John Kanefsky.[22] The database contains a list of all steam engines (more precisely, those for which some historical evidence has been found) erected in Great Britain over the period 1700–1800. We have limited ourselves to the period 1760–1800, as the period before 1760 was entirely dominated by the Newcomen design and thus was characterized by absence of variety and differentiation.[23]

The database contains 1370 engines for the period 1760–1800. Each of these engines is coded as a string of seven alleles that describes the engine design as a point in a seven-dimensional design space. Dimensions and alleles are given in Table 5.2. The design dimensions have been constructed in such a way that each design could be coded as a unique string, thus covering the most relevant dimensions of early steam-engine technology. After having coded each engine in the database as a design string according to the classification of the design space in Table 5.2, we constructed yearly frequency distributions and computed the entropy and mutual information values.

Note that we have considered three-year moving averages of the yearly entropy and mutual information values in order to smooth short-term fluctuations and obtain a ‘neater’ pattern. The results in the figures are shown per year, where each year stands for the in-between year of a three-year period. The transformation of yearly values into three-year moving averages does not in any way affect our conclusions.

4.1.3 Results on steam engines

From the results, it immediately becomes clear that variety (entropy) and differentiation (mutual information) have both increased very rapidly from 1774 onwards when the Watt engine became a popular design next to the

Table 5.2 Design space of steam-engine technology (S = 192)

Steam engine		
Number of observations:		*1370*
Time span:		*1760–1800*
Area:		*Great Britain*
X_1	Pressure	
$A_1 = 2$	0 low, 1 high	
X_2	Condenser	
$A_2 = 2$	0 yes, 1 no	
X_3	Action	
$A_3 = 2$	0 single acting, 1 double acting	
X_4	Compounding	
$A_4 = 2$	0 yes, 1 no	
X_5	Motion	
$A_5 = 3$	0 reciprocating, 1 rotary, 2 water returning	
X_6	Top	
$A_6 = 2$	0 open, 1 closed	
X_7	Cylinder	
$A_7 = 2$	0 single, 1 double	

older Newcomen design (Figure 5.5). The rise in variety and differentiation levelled off around ten years later (more or less from 1785). What is also clear is that, as both entropy and mutual information have been rising, the sum of marginal entropy values must have risen also, following Formula (5.13). This shows that the technological evolution of the steam engine has been characterized by the introduction of new alleles in several dimensions accounting for the rise in the sum of marginal entropy values. The introduction of new alleles has been such that both the variety in designs and the degree of differentiation in design families have risen. Put another way, the design variety has been made possible by the development of new alleles that are combined in highly non-random ways.

Closer inspection of Figure 5.5 also shows that during the 1770s and early 1780s the rise of entropy precedes increases in mutual information. We understand this as probably being due to the fact that new combinations of alleles were tried first, leading to an increase of variety. However, some of these new combinations did not reach adequate levels of fitness, and so we see that, with a delay, mutual information 'catches up' with the

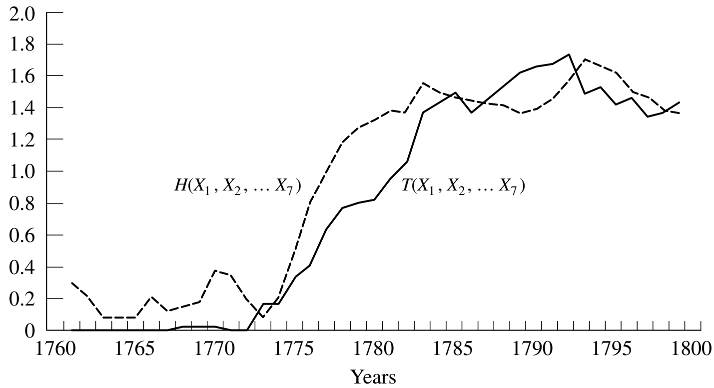


Figure 5.5 N-*dimensional entropy (*H*) and mutual information (*T*) for steam-engine designs*

entropy, which means that the product population is clustering around some specific points of the landscape. In other words, we have first a phase of exploration and discovery of new areas of the landscape, followed by concentration in some points that are likely to be local optima. The 'levelling-off phase' seems to suggest that from the late 1780s a stable pattern of differentiation finally emerged.

Results on two-dimensional mutual information values are depicted in Figure 5.6. The figure shows along which couples of dimensions differentiation has been most pronounced. Hence, these results are also informative about the nature of the technological interdependencies (epistatic relations) among the constituting elements of our design space. The highest mutual information values are reached by the pair $T(X_2,X_6)$, which reflect the interdependence between condensation and the closed-top cylinder. Separate condensation and the closed top cylinder are the two salient features distinguishing Watt type of engines (0100010) from the Newcomen atmospheric engine without condensation and open top (0000000). Importantly, the high values of $T(X_2, X_6)$ are not temporary but continue during the whole period considered. These results thus confirm the thesis of an emergence of a pattern of differentiation.

The other couples of dimensions with high mutual information values are $T(X_2, X_3)$, $T(X_2, X_5)$, $T(X_3, X_5)$, $T(X_3, X_6)$ and $T(X_5, X_6)$. What becomes clear from these results is that the high values are limited to four dimensions: X_2, X_3, X_5 and X_6 (respectively, with/without condenser, single/double action, reciprocating/rotary/water returning, and open/closed top). As explained above, dimensions X_2 and X_6 differentiate Newcomen and Watt

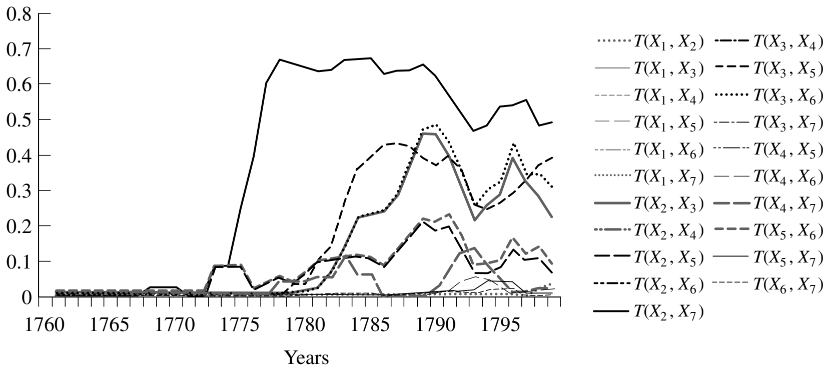


*Figure 5.6 Two-dimensional mutual information (*T*) for steam-engine designs*

engines. Dimensions X_3 and X_5 concern different types of solutions to deliver particular types of motion. Double action was a typical feature of Watt rotary engines (0110110), while Newcomen engines delivering rotary motion made use of either returning a stream over a waterwheel (0000200) or directly by alternatively using two cylinders (0000101).

4.1.4 Discussion

The pattern of growing variety and differentiation of early steam-engine technology suggests that newly developed designs did not simply substitute older designs, but enlarged total design variety. Our analysis shows that the 'linear' view of the early history of the steam engine is essentially untenable. Instead, technological evolution in this period is better characterized by progressive differentiation into distinct design families.[24] In terms of the NK model, the clustering of the product population around some specific designs can be seen as a reflection of designers occupying local optima in the fitness landscape.

The process of differentiation proceeded along four specific technical dimensions (with/without condenser, single/double action, reciprocating/rotary/water returning, and open/closed top). These dimensions may well be related to different user contexts, in particular to the cheapness of coal and to the desired properties of the rotary motion. The pattern of specialization we find contradicts received histories of early steam-engine evolution that point to a process of substitution between Watt engines and Newcomens. Although Watt's inventions are considered to have solved the main shortcomings of the Newcomen engine, it is misleading to assume that they led to the substitution of Newcomen engines.

Regarding the superiority of Watt's fuel efficiency, one can understand the limited substitution of Newcomen engines by Watt engines, taking into account the higher costs of erection and maintenance of the Watt engine. In this respect, von Tunzelmann (1978) has argued that in areas where coal was cheap enough, the Newcomen engine had an important advantage due to its lower costs of installation and maintenance. Besides, whereas the Newcomen engine was well within the engineering capabilities of the time, the Watt engine imposed very compelling requirements on the degree of accuracy of the various components of the engine. This points to the existence of a fundamental trade-off concerning fuel-efficiency versus simplicity of construction and maintenance.[25]

Regarding the type of motion that Watt engines were capable of delivering, the significance of Watt's design modifications also requires further nuance. Although Watt's inventions for supplying rotary motion were highly celebrated (Dickinson and Jenkins 1927), they should not by any means be considered definitive, especially given the accuracy of workmanship of the time. We are aware of many cases of unsatisfactory performance of Watt rotary engines in textile mills.[26] This explains why Watt engines only partially substituted alternative designs that delivered rotary motion.

Interestingly enough, there was an attempt to develop a 'hybrid' engine combining the simplicity of Newcomen with the fuel-efficiency of Watt. This was the 'improved atmospheric engine' patented by Symington in 1787 (0100000). Unfortunately, we have scant information on this engine (especially on its actual fuel-efficiency compared to Watt). We know that about twenty of this type of engine were erected mainly in Scotland and that they generally proved rather successful.[27] Some historians of technology (Dickinson and Jenkins 1927) have dismissed Symington simply as a 'schemer' who tried to circumvent Watt's patent.[28] Our results instead suggest that his attempt to merge the two separate design trajectories of the Newcomen and Watt designs was genuinely aimed at solving a teething trade-off.

To summarize, the existence of various user contexts implied that engine designs be differentiated in order to provide adequate responses to the specific demands of the various user sectors. In our case, this determined a divergence of design trajectories, a process akin to speciation in biology. In a companion paper (Frenken and Nuvolari 2002), we study the pattern of specialization of different type of steam engines in the various user contexts in greater detail using data on the sector of application of engines.

4.2 Aircraft

4.2.1 Aircraft history

Both historians and economists have analysed the development of aircraft technology in considerable detail (for example Miller and Sawers 1968, Constant 1980, Bilstein 1996). Although these studies differ in their perspectives and methodologies, there is a general consensus on the main stages of aircraft development, which can be divided into four periods.

The early history of aircraft from the turn of the century to roughly 1930 is characterized by a large variety of designs and limited demand. A large number of new, small firms experimented with various designs and materials. This period is commonly considered an explorative stage in the industry characterized by a great deal of trial and error. During this period, series production remained limited, causing production costs and prices to be too high for mass consumption.

The second stage, covering the 1930s and early 1940s, has been marked as the period of technological convergence towards what has been termed a 'dominant design' (Abernathy and Utterback 1978). The Douglas DC3 developed in the mid-1930s is generally considered the exemplar of this dominant design. The DC3 is an all-metal, monocoque, piston-propeller monoplane with twin engines placed under the wings. Production costs of this design rapidly fell due to its commercial success in both military and civil aviation. In the early 1940s, total production of the DC3 reached 10000 models (Jane's 1978). The DC3 design also provided the basis of the development of a whole product family developed throughout the 1940s and the 1950s, including the DC4, DC5, DC6 and DC7. At the time, many firms, including Boeing, imitated the DC designs in their piston propeller product lines for passenger aircraft and bombers.

The third stage, covering the period of the 1940s and 1950s, is characterized by the introduction of jet engines. The first experiments with jet engines go back to the Second World War, but their successful application in both military and civil aircraft first took place in the 1950s. The transition from piston-propeller to jet engines has been widely recognized as a technological revolution, which has established a shift in the prevailing 'technological paradigm' (Constant 1980, Dosi 1982). The introduction of jet engines did not simply replace piston-propeller engines in existing designs, but also led to the development of new technologies in other parts of the aircraft, notably the introduction of swept and delta wings that were better able to cope with the increased engine power of jet engines. The revolutionary nature of jet-engine technology can be further supported by the fact that the Douglas, as the most successful company in large piston-propeller aircraft, lost its leading position to Boeing, a company that came to dominate the turbofan passenger aircraft industry.

The fourth stage of aircraft development has been characterized by the further diffusion of jet engines in smaller aircraft, including business aircraft and short-range passenger aircraft. In the period after the 1950s, no major change in aircraft design has taken place as innovative activities increasingly shifted from aircraft design to avionics.

4.2.2 Aircraft data

The data on aircraft design concern the alleles of six design dimensions and covers the period 1913–84. As aircraft development only took off in the early 1900s, the data can be considered to cover the larger part of aircraft history. The choice of the six dimensions and its alleles is based on the limitation posed by the data source, which concerns photographs of aircraft designs. Admittedly, other dimensions that are known to have played an important role, including the types of landing gear and the type of materials used, could not be coded due to the limitation of the source materials. The photographs were drawn from Jane's (1978, 1989) encyclopaedia on aviation, which is known to be among the most comprehensive encyclopaedias of aviation and aircraft designs from all countries. The data of the six dimensions have been compiled for a sample of 731 aircraft models (Table 5.3), corresponding to a sample covering other variables not used here, previously assembled by Paolo Saviotti.[29]

The frequency distributions of designs that are used to measure entropy and mutual information at particular moments in time are not the yearly distributions of product designs. In this case, a year is too short a timespan, as aircraft designs are typically products that remain on offer for many years after their introduction. We used ten-year distributions, but calculations for five-year and 15-year distributions yielded the same trends as discussed below.

The results in the figures below are shown still using a yearly basis, where each year covers a ten-year period. Thus the distribution of designs associated with a specific year corresponds to a time period of ten years beginning in that year. In other words, the year 1913 stands for the distribution of designs introduced between 1913 and 1922; the year 1914 stands for the distribution of product designs introduced between 1914 and 1923, and so on.

4.2.3 Results on aircraft

The results on entropy and mutual information for aircraft are given in Figure 5.7. Entropy increased in the early decades and decreased only slightly in the 1930s. In the 1940s and early 1950s entropy increased rapidly, again to level

Table 5.3 Design space of aeroplane technology (S = 2520)

Aeroplane		
Number of observations:		*731*
Time span:		*1913–84*
Area:		*World*
X_1	Engine type	
$A_1 = 5$	0 piston-propeller, 1 turboprop, 2 jet, 3 turbofan, 4 rocket	
X_2	Number of engines	
$A_2 = 7$	0 one, 1 two, 2 three, 3 four, 4 six, 5 eight, 6 twelve	
X_3	Number of wings	
$A_3 = 3$	0 monoplane, 1 biplane, 2 triplane	
X_4	Wing type	
$A_4 = 4$	0 straight, 1 delta, 2 swept, 3 variable swept	
X_5	Number of tails	
$A_5 = 2$	0 one, 1 two	
X_6	Number of booms	
$A_6 = 3$	0 one, 1 two, 2 three	

off in the late 1950s. Mutual information shows fewer fluctuations, with a general upward trend. Notably, mutual information rose substantially during the period of the 1940s and 1950s and levelled off thereafter. The results suggests that the long-term evolution of aircraft is characterized by both growing variety and growing differentiation into different design families.

The results for the pair-wise mutual information in Figure 5.8 prove informative with respect to the dimensions along which the differentiation process has taken place. It is clear that the rise in mutual information in the post-war period is primarily related to rising mutual information between the engine type and the wing type $T(X_1,X_4)$, between the engine type and the number of engines $T(X_1,X_2)$, and between the number of engines and the wing type $T(X_2,X_4)$. The values for these three pairs of design dimensions have increased very rapidly. The emergence of design families can thus be related to the interdependencies between these design dimensions. The local optima in fitness landscapes are thus primarily characterized by the different alleles engine type, wing type, and the number of engines. Counting the various designs in the final period after 1960 leads us to distinguish between four design families (Frenken 2001): one- and two-engine piston-propeller aircraft with straight wings, two-engine turboprop mono-

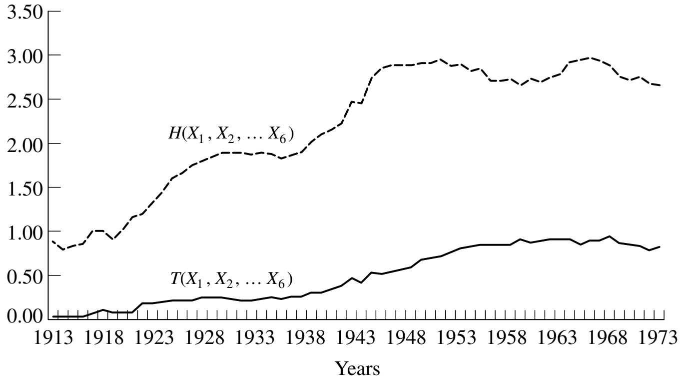


Figure 5.7 N-*dimensional entropy (*H*) and mutual information (*T*) for aircraft designs*

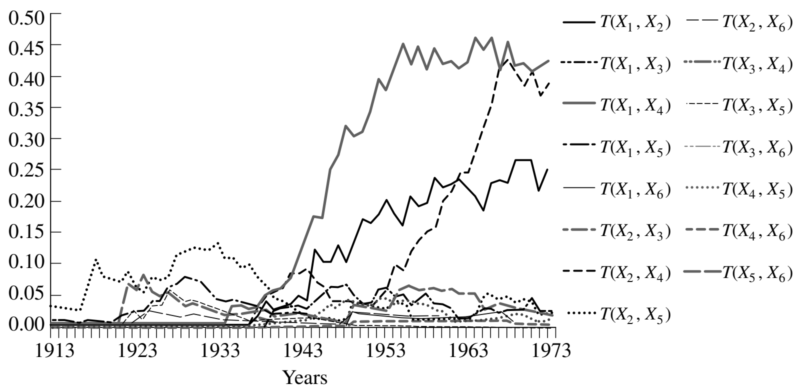


Figure 5.8 *Two-dimensional mutual information (*T*) for aircraft designs*

planes with straight wings, one- and two-engine jet aircraft with delta wings and two-, three- and four-engine turbofan aircraft with swept wings.

Epistatic relations among other pairs of dimensions do not show high dependence, suggesting that dimensions X_3, X_5 and X_6 have not been constitutive for the emergence of design families. All two-dimensional mutual information values including X_3, X_5 or X_6 remain low throughout the period with the exception of the value $T(X_2,X_5)$, This value shows some increase in the 1920s and early 1930s and indicates the common use of an uneven number of engines in two-tail aircraft design, with one engine placed between the tails. After the 1930s, however, two-tail aircraft designs were hardly being used, which shows that this trajectory has proven a dead end.

4.2.4 Discussion

From our results we conclude that the history of aircraft technology is characterized by a progressive development of designs into four distinct families. Though not entirely differing with the histories of the aircraft industry as sketched before, these results offer a number of new insights into its evolutionary dynamics.

First, the emergence of a dominant design in the 1930s commonly associated with the Douglas DC3 had only a limited effect on the total design variety in the industry. The results on aircraft entropy show that the increase in variety was indeed halted during the 1930s, but did not decrease substantially. Second, the advent of jet-engine aircraft in the 1940s and 1950s contributed, as expected, to design variety, with entropy values rapidly rising during this period. However, after the 1950s entropy remained at a high level, suggesting that jet-engine design did not fully substitute propeller designs. Instead, a pattern of differentiation occurred, as indicated by the rising values of mutual information, with piston-propeller and turbo-propeller engine design coexisting alongside jet and turbofan engine designs.

We understand this stable pattern of differentiation as reflecting the different uses of different aircraft designs found in an earlier study that related engine types to market applications (Frenken 2000). Piston-propeller engine design has become dominant in low-cost, small-distance operations including trainer aircraft, business aircraft and agricultural aircraft. Turbo-propeller engine aircraft are used for small-distance passenger aircraft and military transport, while turbofan-engine aircraft are used for medium- and long-distance passenger aircraft. Finally, jet engines are predominantly used in high-speed fighter aircraft.

Note that the history of aircraft technology shows some interesting parallels with early steam-engine technology in that both technologies have witnessed the introduction of a revolutionary design (the jet engine and Watt's engine, respectively). Yet, in both industries the introduction of the revolutionary design has not so much led to a substitution process, but rather to a process of progressive differentiation into different design families.

4.3 Helicopters

4.3.1 Helicopter history

Though the concept of helicopters has a long history that goes back to China in about 400 BC, the first successful helicopter dates back to 1939 with the development of the VS-300 by Sikorsky (Taylor 1995). The advent of helicopter technology quickly received interest from armies and navies, because of helicopters' capacity to evacuate people from areas that were not accessible by aeroplanes. The military demand for helicopters induced a great deal of explorative activity in the 1940s and 1950s, including variations in the type of engine, the number of rotors, and the number of blades. At the time, commercial expectations were high, as evidenced by popular magazines predicting that American households would soon have a family of helicopters in the garage.

In the late 1950s, the explorative stage of technological development largely came to an end as design convergence took place with the apparent superior engine performance of turbines to piston engines. According to Bilstein (1996; p. 91), the single-rotor twin-turboshaft Kaman model introduced in 1954 can in hindsight be considered a 'pioneering' design. Hereafter, the twin-engine turboshaft design with one rotor became the 'dominant design'.

Commercially, however, helicopters never became a mass-produced product. Compared to aircraft, the costs and limited range of helicopters impede their wider diffusion in segments currently dominated by conventional aircraft (Taylor 1995). Instead, most helicopters are used for transporting people in areas not accessible by aircraft (such as military troops or offshore oil-platform personnel), while niche applications exist for a variety of uses, including ambulance operations and fighter operations.

4.3.2 Helicopter data

The data on helicopters concern the alleles of five design dimensions and cover the period 1940–83. The data of the five dimensions have been compiled for a sample of 144 helicopter models (Table 5.4). As for the data on aircraft, the helicopter data have been compiled on the basis of observable characteristics on photographs and correspond to the sample previously compiled by Paolo Saviotti[30] from Jane's (1978, 1989) encyclopaedia on aviation.

As for aircraft, the frequency distributions of designs that are used to measure entropy and mutual information at particular moments in time are not the yearly distributions of designs. We used again ten-year distributions, but the calculations for five-year and 15-year distributions yielded the same trends as in the results based on ten-year distributions discussed below. The results in the figures are shown per year, where each year stands for the first year of a ten-year period.

4.3.3 Results on helicopters

The results on entropy and mutual information are given in Figure 5.9. Interestingly, the results on helicopter variety and differentiation show

Table 5.4 Design space of helicopter technology (S = 420)

Helicopters	
Number of observations:	*144*
Time span:	*1940–83*
Area:	*World*
X_1	Engine type
$A_1 = 5$	0 piston, 1 piston turbo, 2 ramjet, 3 gas generator, 4 turboshaft
X_2	Number of engines
$A_2 = 3$	0 one, 1 two, 2 three
X_3	Number of blades
$A_3 = 7$	0 two, 1 three, 2 four, 3 five, 4 six, 5 seven, 6 eight
X_4	Number of shafts
$A_4 = 2$	0 one, 1 two
X_5	Number of rotors per shaft
$A_5 = 2$	0 one, 1 two

patterns that are altogether different from the results on early steam engines and aircraft. After a short period of rising values, entropy has fallen from 1955 onwards, showing that product variety in product designs has also fallen. The value for mutual information peaked earlier in 1949 and thereafter also shows a declining trend. Note that the decline in mutual information has been relatively greater than the decline in entropy values (mutual information halved during the period 1950–80). This suggests that the variety that remained was increasingly based on small variants around a single dominant design, which is the one-rotor turboshaft helicopter covering the large majority of models made from the 1950s onwards.

The two-dimensional mutual information values for helicopters also show decreasing trends (Figure 5.10). Only one pair of dimensions $T(X_2,X_3)$ shows the highest values over the whole period, reflecting complementarities between the number of engines and the number of blades. This relationship points to the common use of more blades when more engines are incorporated in a helicopter design to carry the higher weight. These variations remained within the dominant family of one-rotor turboshaft helicopters.

4.3.4 Discussion

The fall in mutual information accompanied by a fall in entropy suggests that after a brief period of differentiation, we have a prolonged phase in

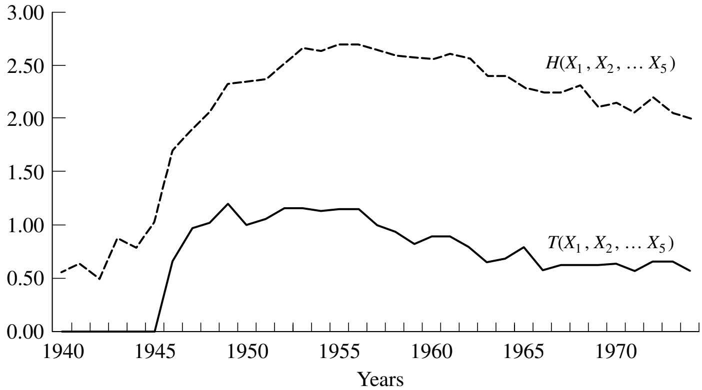


Figure 5.9 N-*dimensional entropy (*H*) and mutual information (*T*) for helicopter designs*

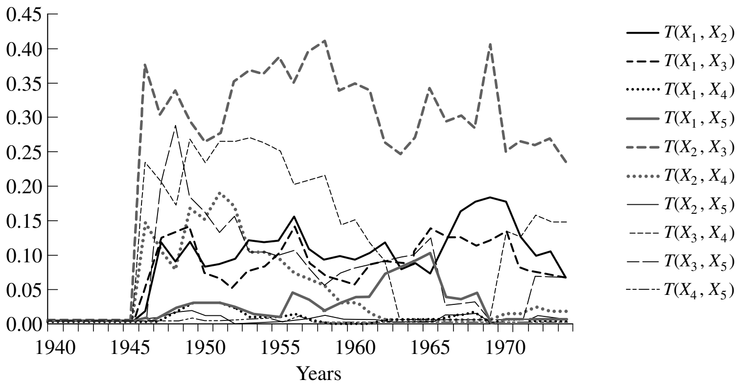


Figure 5.10 *Two-dimensional mutual information (*T*) for helicopter designs*

which design variety decreases. The results correspond to Bilstein's (1996, p. 91) historical account that identified the single-rotor twin-turboshaft design as the dominant design emerging in the 1950s. Our analysis is also in line with findings by Saviotti and Trickett (1992, p. 116), who found that the single-rotor turboshaft helicopters increased their share in the population from around 30 per cent in the late 1950s to around 80 per cent in the early 1980s.

In the case of helicopter technology, de-differentiation cannot be attributed to absence of heterogeneity in demand. In fact, Saviotti and Trickett (1992) distinguish between up to 22 different uses of helicopters, ranging from fighter operations to military transport to ambulance to business transport. User heterogeneity may well be at least as high as in aircraft industry, even though sales in the helicopter industry are only a fraction of those in the aircraft industry. Given the heterogeneity of helicopter demand and the process of de-differentiation of helicopter supply, Saviotti and Trickett conclude that heterogeneity in demand is met by modular designs capable of being used in a variety of user contexts. In this context, one must think of helicopters in which the interior is easily adapted without changing the helicopter design itself.

The results still leave open the question as to why heterogeneity in user contexts in early steam-engine and aircraft design have triggered differentiation, while heterogeneity in helicopters users has not led to a sustained pattern of differentiation. Following Frenken et al. (1999b), one can explain the de-differentiation in helicopter technology by the existence of competition between helicopter technology and aircraft technology. Within the market for air transport, helicopter technology itself operates within a relatively small niche, which is bounded by the presence of aircraft technology. Over the past few decades, single-rotor helicopter performance has been limited by a flight range of around 1000 km, a speed of 300 km/h and payload of around 10000 kg. The halt in improvements does not reflect technical difficulties, but competition with aircraft: further improvements in speed, range or payload are technically perfectly realizable, but would lead helicopters to compete with small cheap aircraft covering the market segments of longer distances, higher speeds and higher payload. Put another way, the range of performance levels that helicopters are technically capable of reaching has not been fully explored due to the presence of cheaper aircraft technology.

5. CONCLUSION

We started our study by introducing the NK model as a formal model of complex evolving systems that are characterized by interdependencies among their constituting components. We proposed a number of generalizations to the original NK model to account for the specificities of technological evolution. By examining the properties of this generalized NK model, we concluded that technological development in complex technologies is likely to lead to a process of differentiation of designs into distinct families. This view contradicts models of technological substitution that depict competition among designs as a one-dimensional (cost-based) process that leaves room for only one surviving technology.

To analyse the evolutionary pattern of technological development in terms of changes in variety and differentiation, we proposed the methodology of entropy statistics. Entropy provides us with a comprehensive measure of design variety, while mutual information indicates to what extent this variety is non-random, that is, clustered in specific areas of the design space. The existence of multiple clusters indicates the presence of local optima in the technology's fitness landscape.

We applied the entropy statistics to data on design dimensions of three technologies. The results confirmed our hypothesis of increasing variety through differentiation for aircraft and steam engines, while the de-differentiation process of helicopter technology could be attributed to the presence of competing aircraft models. Furthermore, the empirical results offered us insights into the (quantitative) evolution that differ from the received histories of steam engines and aircraft. We found that the evolution of two technologies is better described as an evolutionary process of differentiation than as a linear substitution process. Obviously, a next step is to apply the methodology presented in this chapter to other technologies. The proposed methodology can be applied to any technology, given that sufficient empirical data are available on the relevant design dimensions of the technology in question.

APPENDIX: DERIVATION OF MUTUAL INFORMATION FOR ZERO ENTROPY AND MAXIMUM ENTROPY

Entropy is zero when one design occurs with frequency one, implying that the alleles incorporated in this design also occur with frequency one. Therefore, the sum of marginal entropy values equals zero, implying that mutual information equals zero:

$$T(X_1, \ldots, X_N) = \left(\sum_{i=1}^{N} H(X_i) \right) - H(X_1, \ldots, X_N)$$

$$T(X_1, \ldots, X_N) = \left(\sum_{i=1}^{N} -1 \cdot \ln 1 \right) - -1 \cdot \ln 1 = 0 + 0 = 0$$

Entropy is maximum when all possible designs in design space have an equal frequency $1/S$. In that case, the alleles along each dimension also have an equal frequency with marginal frequencies equalling $1/A_i$. Mutual information becomes:

$$T(X_1, \ldots, X_N) = \left(\sum_{i=1}^{N} H(X_i) \right) - H(X_1, \ldots, X_N)$$

$$T(X_1, \ldots, X_N) = \left(\sum_{i=1}^{N} \ln A_i \right) - \ln S$$

$$T(X_1, \ldots, X_N) = \ln \left(\prod_{i=1}^{N} A_i \right) - \ln S$$

$$T(X_1, \ldots, X_N) = \ln S - \ln S = 0.$$

NOTES

* We are grateful to John Kanefsky for generously providing us with the updated version of his dataset on British steam engines, and to Paolo Saviotti for the use of his data on aircraft and helicopters previously collected for a project funded by the ESCR. We thank Nick von Tunzelmann for helpful discussions. An earlier draft of this chapter was presented at the Second European Meeting on Applied Evolutionary Economics in Vienna, September 2001. We thank the participants to the workshop and, especially, Arnulf Grübler for their comments.

1. Bradshaw (1992) uses the concepts of Simon to provide an interesting account of the Wrights' development of early aircraft technology.
2. See Kauffman and Macready (1995), Levinthal (1997), Frenken et al. (1999a), Auerswald et al. (2000), Gavetti and Levinthal (2000), Kauffman et al. (2000), Marengo et al. (2000), Rivkin (2000), Valente (2000), Fleming (2001), Fleming and Sorenson (2001), Frenken (2001). See also the discussion by Simon (2002) on the relationship between Kauffman's NK model and Simon's (1969 [1996]) early work.
3. Notable expectations are Fleming (2001) and Fleming and Sorenson (2001).
4. See also David (1975) in his discussion on localized technological change.
5. The combinatorial nature of the design space of a system requires that dimensions are orthogonal to one another. Therefore, one dimension of a system cannot correspond with an allele of another dimension in the same system. For example, the description of alleles of the engine dimension as gasoline ('0'), electric ('1') and steam ('2') implies that the type of battery used in electric engines cannot count as another dimension in the description of the vehicle as a system. The choice of a type of battery only constitutes a dimension for electric vehicles, and not for vehicle technologies in general.
6. Note that, since the first allele is labelled '0', the description of alleles of an element ranges from 0 to $A_i - 1$, while the number of alleles ranges from 1 to A_i.
7. A system's architecture has also been termed the system's internal structure (Simon 1969 [1996], Saviotti 1996).
8. The *K* value is an indicator of the complexity of a system's architecture and does not exactly coincide with the system's computational complexity, which can be expressed as the computational time that is required to globally optimize a complex system. On this, see Frenken et al. (1999a).
9. Allowing for mutation in several dimensions at the same time would permit a designer to escape local optima. However, the more dimensions that are allowed to be mutated at the same time, the higher the search costs involved as the number of possible moves increases exponentially with the number of dimensions that is allowed to be mutated at the same time. One can thus argue that designers are expected to search in only a few dimensions at the same time. On this issue, see Frenken et al. (1999a) and Kauffman et al. (2000).

10. For more properties of the NK model, see Kauffman (1993), Altenberg (1997), and Frenken et al. (1999a).
11. This perspective on fitness differs from the NK model applied to process technology where fitness is expressed only by a single cost criterion (Auerswald et al. 2000, Kauffman et al. 2000).
12. Altenberg's generalized NK model also allows one to model search by adding new components to a system, increasing N while keeping the number of selection criteria F constant. On this, see Altenberg (1994, 1995).
13. This is a relatively simple function sometimes applied in multi-criteria analysis of project selection (Nijkamp et al. 1990). This function implies that a loss in fitness of one function can infinitely be substituted by an increase in other functions. Various alternative functions exist to derive a fitness value or 'utility' from a collection of characteristics (Lancaster 1966, 1979).
14. Compare the SCOT approach of Pinch and Bijker (1984), who stress the interpretative flexibility of the meaning and use of artefacts. Here, the sociology of technology meets evolutionary economics.
15. Note that in the case of heterogeneous user groups, some weights can equal zero, while in the case of a homogeneous user group in formula 5.7, all weights are by definition positive. A zero weight in a homogeneous user population would imply that the feature does not count as a function for anyone.
16. In information theory entropy is computed using the logarithm of two instead of the natural logarithm taken here (Theil 1967, 1972, Frenken et al. 1999b).
17. $0 \cdot \ln(0) \equiv 0$.
18. A more elaborated account can be found in Frenken and Nuvolari (2002).
19. In this respect, Dickinson (1938) can be considered an exemplary reference.
20. A number of Newcomen engines were successfully used to raise water over a water wheel which, in turn, delivered rotary motion for factory machinery. These types of engine were usually called returning engines.
21. Von Tunzelmann (1978, p. 263).
22. For more details on the original data see Kanefsky (1979). For a more accessible reference, see Kanefsky and Robey (1980).
23. To be more precise, apart from the Newcomen design a second engine design was available before 1760. This design is the Savery engine, which we have excluded altogether from the analysis as it did not meet the classification of our design space. We consider the Savery engine to be a steam pump rather than a steam engine as it lacks the characteristic piston–cylinder arrangement characteristic of all the other steam engines. The exclusion of the Savery engine should not affect our results since only 33 Savery engines are present in the original data. More details on the Savery engine can be found in Frenken and Nuvolari (2002).
24. A similar conclusion based on historical grounds, stressing the role of variety, has been reached by Von Tunzelmann (1978, p. 24): 'It is misleading to see the pattern of progress [in steam-engine technology] as linear and inevitable: in explaining the direction and the chronology of "technical progress" in the economist's sense, it is vital to keep this diversity in mind.'
25. Joseph Bramah stated that the Newcomen engine had over Watt 'an infinite superiority in terms of simplicity and expense'. John Smeaton, one of the leading engineers of the time, considered that the Watt engine demanded too high standards for construction and maintenance. See Harvey and Downs-Rose (1980, pp. 22–3).
26. See Hills (1970, pp. 179–86). Many contemporary engineers believed that the rotary drive produced by a water-returning engine was much more regular and, in the end, 'better' than the one obtained from a rotary Watt engine. See also von Tunzelmann (1978, pp. 142–3).
27. On the Symington engine see Harvey and Downs-Rose (1980, ch. 3).
28. Dickinson and Jenkins (1927, p. 318). See also Farey (1827, p. 656).
29. See Saviotti and Bowman (1984) and Saviotti (1996).
30. See Saviotti and Trickett (1992) and Saviotti (1996).

REFERENCES

Abernathy, W.J. and Utterback, J. (1978), 'Patterns of industrial innovation', *Technology Review*, **50**, 41–7.
Altenberg, L. (1994), 'Evolving better representations through selective genome growth', in *Proceedings of the IEEE World Congress on Computational Intelligence, Part 1*, Piscataway, NJ: IEEE, pp. 182–7.
Altenberg, L. (1995), 'Genome growth and the evolution of the genotype-phenotype map', in W. Banzhaf and F.H. Eckman (eds), *Evolution and Biocomputation*, Berlin and Heidelberg: Springer, pp. 205–59.
Altenberg, L. (1997), 'NK fitness landscapes', in T. Back, D. Fogel and Z. Michalewicz (eds), *The Handbook of Evolutionary Computation*, Oxford: Oxford University Press.
Arthur, W.B. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *The Economic Journal*, **99**, 116–31.
Auerswald, P., Kauffman, S., Lobo, J. and Shell, K. (2000), 'The production recipes approach to modeling technological innovation: an application to learning-by-doing', *Journal of Economic Dynamics and Control*, **24**, 389–450.
Basalla, G. (1988), *The Evolution of Technology*, Cambridge: Cambridge University Press.
Bilstein, R.E. (1996), *The American Aerospace Industry*, New York: Twayne Publishers / Prentice-Hall.
Bradshaw, G. (1992), 'The airplane and the logic of invention', in R. Giere (ed.), *Cognitive Models of Science*, Minneapolis: University of Minnesota Press.
Constant II, E.W. (1980), *The Origins of the Turbojet Revolution*, Baltimore, MD: Johns Hopkins University Press.
David, P.A. (1975), *Technical Choice, Innovation and Economic Growth. Essays on American and British Experience in the Nineteenth Century*, London: Cambridge University Press.
Dickinson, H.W. (1938), *A Short History of the Steam Engine*, London: Cambridge University Press.
Dickinson, H.W. and Jenkins, R. (1927), *James Watt and the Steam Engine*, London: Encore.
Dosi, G. (1982), 'Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change', *Research Policy*, **11**, 147–62.
Farey, J. (1827), *A Treatise on the Steam Engine* (reprint, Newton Abbot: David & Charles, 1971).
Fleming, L. (2001), 'Recombinant uncertainty in technological search', *Management Science*, **47**, 117–32.
Fleming, L. and Sorenson, O. (2001), 'Technology as a complex adaptive system: evidence from patent data', *Research Policy*, **30**, 1019–39.
Freeman, C. (1991), 'Innovation, changes of techno-economic paradigm and biological analogies in economics', *Revue Economique*, **42**, 211–32.
Frenken, K. (2000), 'A complexity approach to innovation networks', *Research Policy*, **29**, 257–72.
Frenken, K. (2001), *Understanding Product Innovation using Complex Systems Theory*, PhD thesis, University of Amsterdam and University of Grenoble.
Frenken, K. and Nuvolari, A. (2002), 'The early development of the steam engine: an evolutionary interpretation using complexity theory', *Industrial and Corporate Change* **13**, 419–50.

Frenken, K., Marengo, L. and Valente, M. (1999a), 'Interdependencies, near-decomposability and adaptation', in T. Brenner (ed.), *Computational Techniques for Modelling Learning in Economics*, Boston: Kluwer.

Frenken, K., Saviotti, P.P. and Trommetter, M. (1999b), 'Variety and niche creation in aircraft, helicopters, motorcycles and microcomputers', *Research Policy*, **28**, 469–88.

Gavetti, G. and Levinthal, D. (2000), 'Looking forward and looking backward: cognitive and experiential search', *Administrative Science Quarterly*, **45**, 113–37.

Harvey, W. and Downs-Rose, G. (1980), *William Symington: Inventor and Engine Builder*, London: Northgate.

Henderson, R. and Clark, K. (1990), 'Architectural innovation', *Administrative Science Quarterly*, **35**, 9–30.

Hills, R. (1970), *Power in the Industrial Revolution*, Manchester: Manchester University Press.

Jane's (1978), *Jane's Encyclopedia of Aviation*, London: Jane's Publishing.

Jane's (1989), *Jane's Encyclopedia of Aviation*, London: Studio Editions.

Kanefsky, J.W. (1979), *The Diffusion of Power Technology in British Industry 1760–1870*, PhD thesis, University of Essex.

Kanefsky, J.W. and Robey, J. (1980), 'Steam engines in 18th-century Britain: a quantitative assessment', *Technology and Culture*, **21**, 161–86.

Kauffman, S.A. (1993), *The Origins of Order. Self-Organization and Selection in Evolution*, Oxford: Oxford University Press.

Kauffman, S.A. and Levin, S. (1987), 'Towards a general theory of adaptive walks on rugged landscapes', *Journal of Theoretical Biology*, **128**, 11–45.

Kauffman, S.A. and Macready, W.G. (1995), 'Technological evolution and adaptive organizations', *Complexity*, **1**, 26–43.

Kauffman, S.A., Lobo, J. and Macready, W.G. (2000), 'Optimal search on a technology landscape', *Journal of Economic Behavior and Organization*, **43**, 141–66.

Lancaster, K.J. (1966), 'A new approach to consumer theory', *Journal of Political Economy*, **14**, 133–56.

Lancaster, K.J. (1979), *Variety, Equity and Efficiency*, New York: Columbia University Press.

Langton, C.G. (1990), 'Computation at the edge of chaos', *Physica D*, **42**, 12–37.

Levinthal, D.A. (1997), 'Adaptation on rugged landscapes', *Management Science*, **43**, 934–50.

Levinthal, D.A. (1998), 'The slow pace of rapid technological change: Gradualism and punctuation in technological change', *Industrial and Corporate Change*, **7**, 217–47.

Marengo, L., Dosi, G., Legrenzi, P. and Pasquali, C. (2000), 'The structure of problem-solving knowledge and the structure of organizations', *Industrial and Corporate Change*, **9**, 757–88.

Miller, R. and Sawers, D. (1968), *The Technical Development of Modern Aviation*, London: Routledge & Kegan Paul.

Mokyr, J. (1990), *The Lever of Riches*, Cambridge: Cambridge University Press.

Nelson, R.R. (1995), 'Recent evolutionary theorizing on economic change', *Journal of Economic Literature*, **33**, 48–90.

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

Nijkamp, P. Rietveld, P. and Voogd, H. (1990), *Multicriteria Evaluation and Physical Planning*, Amsterdam: North-Holland.

Pinch, T.J. and Bijker, W.E. (1984), 'The social construction of facts and artifacts: or how the sociology of science and the sociology of technology might benefit each other', *Social Studies of Science*, **14**, 399–442.

Prigogine, I. and Stengers, I. (1984), *Order out of Chaos*, New York: Bantam.

Rivkin, J.W. (2000), 'Imitation of complex strategies', *Management Science*, **46**, 824–44.

Rosenberg, N. (1976), *Perspectives on Technology*, Cambridge: Cambridge University Press.

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

Sahal, D. (1985), 'Technological guideposts and innovation avenues', *Research Policy*, **14**, 61–82.

Saviotti, P.P. (1996), *Technological Evolution, Variety and the Economy*, Cheltenham, UK and Brookfield, USA: Edward Elgar.

Saviotti, P.P. and Bowman, A. (1984), 'Indicators of output of technology', in M. Gibbons, P. Gummett and B.M. Udgaonkar (eds), *Science and Technology Policy in the 1980s and Beyond*, London: Longman.

Saviotti, P.P. and Metcalfe, J.S. (1984), 'A theoretical approach to the construction of technological output indicators', *Research Policy*, **13**, 141–51.

Saviotti, P.P. and Trickett, A. (1992), 'The evolution of helicopter technology, 1940–1986', *Economics of Innovation and New Technology*, **2**, 111–30.

Simon, H.A. (1969), *The Sciences of the Artificial*, 3rd edn 1996, Cambridge, MA: MIT Press.

Simon, H.A. (2002), 'Near decomposability and the speed of evolution', *Industrial and Corporate Change*, **11**, 597–9.

Taylor, J.W.R. (1995), 'Constraints impeding the commercial use of helicopters', in W.M. Leary (ed.), *From Airships to Airbus. The History of Civil and Commercial Aviation. Volume 1: Infrastructure and Environment*, Washington, DC: Smithsonian Institution Press.

Theil, H. (1967), *Economics and Information Theory*, Amsterdam: North-Holland.

Theil, H. (1972), *Statistical Decomposition Analysis*, Amsterdam: North-Holland.

Theil, H. and Fiebig, D.G. (1984), *Exploiting Continuity. Maximum Entropy Estimation of Continuous Distributions*, Cambridge, MA: Ballinger.

Valente, M. (2000), *Evolutionary Economics and Computer Simulations. A model for the evolution of markets*. PhD thesis, Aalborg University.

Vincenti, W.G. (1990), *What Engineers Know and How They Know It. Analytical Studies from Aeronautical History*, Baltimore, MD: Johns Hopkins University Press.

Von Tunzelmann, G.N. (1978), *Steam Power and British Industrialization to 1860*, Oxford: Clarendon Press.

Ziman, J. (2000), (ed.), *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press.

6. Complementarity constraints and induced innovation: some evidence from the first IT regime

Andreas Reinstaller and Werner Hölzl

1. INTRODUCTION

Schumpeter (1939) distinguished three stages in the process of technical change: (i) invention, that is, the act of creation of a new technology, (ii) innovation, its commercial introduction, and (iii) diffusion, its gradual adoption. Evolutionary economists recognized the importance of Schumpeter's trichotomy, but in the past their work has mostly focused on the last two stages of the process. The inducements and focusing devices leading entrepreneurs to produce new combinations are not analysed in an appropriate way. Technological search is often depicted as random. Neoclassical work on technical change has long studied John Hicks's induced innovation hypothesis in the framework of aggregate production functions. The key insight is that 'a change in the relative prices of the factors of production is itself a spur to invention, and to invention of a particular kind directed to economizing the use of a factor which has become relatively expensive' (Hicks 1932, pp. 124–5). This literature studies the inducement mechanism relying on the principle that a rise in real wages will trigger labour-saving innovation. The problem arising from this type of work is that the aggregate production function framework seems not to be appropriate, as technological change is an inherently microeconomic phenomenon. Second, neoclassical production functions of the Cobb–Douglas, CES or translog type are strongly separable. Separability amounts to the claim that the marginal rate of substitution of any pair of inputs is unaffected by changes in the level of another input.[1] Inputs or groups of inputs cannot be complementary. As the innovation process is not only a microeconomic phenomenon but also determined by the systemic character of firms and the technology they use, this assumption is quite strong.

The aim of this chapter is to analyse how recombinant search is triggered, how it is done and how initial conditions influence the final design of the technological artefacts resulting from this process. We argue that complementarities (non-separabilities) play an important role as focusing devices guiding the search for new combinations. Our analysis takes the perspective of technology adopters and not that of inventors or innovators of new products (Section 2). We illustrate the process of decomposition and recomposition in the presence of binding complementarity constraints with a historical case study on the establishment of the first IT regime at the turn of the nineteenth century (Section 3), and conclude with a discussion of our results (Section 4).

2. COMPLEMENTARITY CONSTRAINTS AND THE DIRECTION OF TECHNICAL CHANGE: THE TRANSITION TOWARDS A NEW TECHNOLOGICAL PARADIGM

The viability of a system depends on how well its different elements fit together. Complementarity refers to the relationship between these elements, that is, how they mutually influence performance. An example of a technical system exhibiting complementarities is the modular system of the personal computer (PC): the choice of best components (processor, motherboard, graphic adapter, software, and so on) does not necessarily imply that this computer works better than another, which consists of syntonic but 'inferior' components. The PC consisting of 'best components' may even not work, for example, if the 'best' processor cannot be put on the 'best' motherboard. Firms can also be seen as entities that channel inputs into a complex organizational structure, characterized by the technology in use. Firms bundle activities into a productive entity by interconnecting them. Some of the links in the resulting network can be very strong, while others may be weaker. These linkages and their strength can result from strict technical (or static) complementarities between elements in this web, but also from tacit or explicit dynamic complementarities that capture learning spillovers. They exist between inputs of an activity, between activities of a firm and between firms.

Complementarities imply that elements in a system are not separable. Taking the personal computer metaphor as example, complementarities imply that a PC may work less well or better with one or another hard-disk controller, but it will not function without one. The PC as a system is thus non-separable from this component. The implications of non-separabilities can be illustrated by Stuart Kauffman's NK model (Kauffman 1993), which provides an easy way of thinking about complementarity interrelationships in systems. This model has been used in recent times by several contribu-

	I_1	I_2	I_3
C_1	X		X
C_2	X	X	
C_3		X	X

(a)

	I_1	I_2	I_3
C_1	X		
C_2	X	X	
C_3	X		X

(b)

Figure 6.1 (a) Symmetrical linkages, N *=3,* K *=1; (b) Asymmetrical version*

tors to evolutionary economics to study the way economic agents explore a fitness landscape of technologies or organizational set-ups in the presence of complex feedback mechanisms (see, for example, Frenken et al. 1999). The NK model simulates the evolution of complex systems in which the elements function interdependently. A system is described as a set of N elements each of which can take on A_i possible values. The number of all possible strings among system elements is called the possibility space S of a system. Its size is given by the Cartesian product of the possible values: $S = A_1 \times A_2 \times \ldots \times A_N$. If a system consists of three binary elements ($N=3$; $A_i=2$), the size of S is equal to $S=A_i^N=2^3=8$. This could be interpreted as a production technology with three inputs or activities, which may take two states: 0 if a productivity-enhancing innovation for this input is absent or 1 when it is present. S would be the space of technological designs. The fitness values of the eight possible designs are taken to represent a productivity measure. Each input makes a contribution to the total productivity of a design. K denotes how interdependent the inputs are. Two extreme cases can be contrasted: if $K=0$, each input's productivity contribution is independent of all others. In this case one global optimum exists and it depicts the case of a perfectly separable neoclassical production function. At the other extreme we have maximum complexity ($K=N-1$), where each input's productivity contribution depends on all other elements. The number of possible optima in this landscape increases with N and K. The number of optima is a measure of the complexity of a system. Very complex ones have many possible 'good', that is, locally optimal, designs. The interrelations need not be symmetric. Matrices (a) and (b) in Figure 6.1 show this in the generalized NK model of Altenberg (1994, 1997). Matrix (a) is symmetrical ($N=3$, K $=1$), as every input is linked to another. Matrix (b) instead shows the asymmetric case, where only input I_1 influences the two others. The C_x indicate technical (Lancastrian) characteristics of each of the inputs. The matrices in Figure 6.1 map technology or inputs into technical characteristics. In terms of the approach of Saviotti and Metcalfe

	I_1	I_2	I_3	I_4
C_1	X		X	
C_2	X	X		X
C_3	X	X	X	
C_4		X		X

(c)

	I_1	I_2	I_3	I_4	I_5	I_6	I_7
C_1	X					X	
C_2		X			X		X
C_3			X			X	X
C_4				X	X		

(d)

Figure 6.2 *(c) Asymmetrical linkages in a generalized NK model with two core inputs* I_1 *and* I_2*; (d) Decomposed version with perfectly modular core inputs (*I_1*–*I_4*) and new coordination or interface inputs (*I_5*–*I_7*)*

(1984), the inputs could be defined as process technology and the technical characteristics these produce as product technology. This shows that a characteristic C_x may be influenced by several activities, while one activity in turn may influence several of these output characteristics. So I_1 produces output characteristic C_1 and influences the output characteristics C_2 and C_3 produced by activities I_2 and I_3 respectively, while output characteristic C_1 results from the joint action of processes I_1 and I_3.

In a recent paper Frenken (2001) argued that activities mediating the interactions of many others constitute the 'core' of an evolving technological paradigm. The more complex the process technology, the higher the likelihood that a change in one component may conflict with the overall performance. This implies that elements in the core are more likely to be changed only rarely so that the performance improvements in a technological paradigm take place mostly by adding or changing inputs that are peripheral, such as inputs I_2 and I_3 in matrices (c) and (d) in Figure 6.2. Even though the system may be more frail with regard to changes in its core, which also restricts the set of profitable moves in the possibility space, it also increases the probability of evolution towards a stable and profitable set-up and reduces technological uncertainty (Caminati 1999). The core represents a design of a process technology that works and that can be further explored.

In a stable economic environment entrepreneurs have an incentive to keep the core elements of their operating technique as they are. If economic conditions change, through events that punctuate the equilibrium, for example shifts in customer preferences or sudden and persistent rises in the prices of some inputs, the strong complementarities in the core may turn into binding constraints. The complementary relations may cause imbalances between related technologies. While under stable conditions complementarities enable the search for new variations of working designs under radical change, they hinder the exploration of new designs. There is an incentive for firms to break up the constraints of complementarities between the different elements of the core and increase its degrees of freedom. Complementarities become powerful focusing devices for technical change. The outcome of the ensuing process of technological search is likely to give rise to radically new solutions, which may entail a complete break-up and reconfiguration of the core.

Schumpeter (1947) described the entrepreneur as the architect of new combinations. Recombining is the ability to recognize linkages between technologies that lead to new products, new output characteristics or increased productivity. If the technological system is complex, it is a non-trivial matter to infer the properties of the whole. It is unlikely that innovators know about the existence and entity of all linkages. Simon (1996) suggested with the Tempus and Hora parable that the best strategy to manage complex problems is to reduce the number of distinct elements in the system by grouping them into subsystems through the suppression of less important relations. The transformation of a complex system into a nearly decomposable one is a process of problem decomposition. The linkages between the single parts and the laws of interactions have to be understood and codified before any process of recombination can start. Von Hippel (1990) emphasized the importance of task-partitioning as an important feature of the innovation process. Cowan and Foray (1997) suggested that problem decomposition was a slow and stepwise process of codification, which involves the modelling of knowledge. The recursive application of the subdivision of tasks leads to an ever better understanding of the relationship between different skills, routines and heuristics, reducing their tacitness or leading to a better understanding of the source that generates strict complementarities. Decomposition is thus a slow process of identification and selection of links to modify the economic performance of artefacts and activities. In this process past experiences are important starting points. For example, as we will see in the case study that follows, the experience American engineers gained in redesigning products to use standardized and interchangeable parts was prior procedural knowledge in the efforts to reorganize the administration.

Task-partitioning and codification are the first steps towards the emergence of a new technical and/or organizational architecture. The process of recomposition is a co-evolutionary process in which managers searching for new process technologies and suppliers of new technological artefacts interact. In this interaction technical change takes place at the level of activities and the level of artefacts. The role of technology suppliers is to provide a variety of new technological artefacts, which the adopters will explore. The new machines embody some of the knowledge codified during the process of decomposition. This can happen when the division of labour has progressed so far that skills and tacit knowledge related to some original inputs are reduced to chunks of structured information. Then a complete codification and embodiment into specific control mechanisms of machines become feasible. Technological closure on the artefacts and dominant designs emerges as soon as a well-defined domain of application is established; that is, the process of recomposition settles to some defined architecture of the process technology. There is a considerable amount of literature, starting with Tushman and Anderson (1986), that has studied this process in detail. Several mechanisms are at work in the process of recomposition. Adapting the work of Wagner and Altenberg (1996) to the economic context, we identify two main influences guiding the act of searching for new process technology designs on the part of entrepreneurs: first, they will try to isolate and substitute those activities which in the act of decomposition have been identified as causes for imbalances in the process technology. Second, they will try to avoid negative influences of the new technology on other parts of their core activities, in order not to lose out to their competitors; that is, they will try to stabilize the function of all other parts of the process technology. The particular skill of the designer of the new process technology lies in selecting linkages that modify the product technology in such a way as to eliminate the causes that have triggered the process of technological search but keep it consistent with the needs of the markets in which the firm operates. Wagner and Altenberg (1996) suggest two mechanisms for regrouping linkages. The first is called *parcellation* and consists of the suppression of interconnections of lower importance between activities influencing the same group of characteristics of the product technology. The second is *integration*, where linkages are selectively established between previously unrelated parts – new and old – of the process technology. Both favour the development of a more modular design of the process technology.

If these mechanisms are also in place in the process of technological search and adoption, then one should first expect a differentiation of existing capital goods and work profiles to the more restricted scope of new activity groups and the integration of new capital goods and work profiles. This entails a process of capital deepening, meaning that the quality of the capital stock is improved in terms of its effects on labour productivity. Second, the regrouping of activities should make process technology more modular. A more modular and less integrated structure at lower levels of the process technology requires integration at a higher level. This implies that the hierarchical structure should deepen. Theoretical results from the generalized NK model support this view because they show that the likelihood of successful improvements increases in modular technology-characteristic maps, as they are less likely to be offset by negative feedbacks (Altenberg 1994). Modularity increases the evolvability of the system as illustrated in Figure 6.2: matrix (c) shows a system with four inputs producing output characteristics C_1 to C_4. Inputs I_1 and I_2 are highly interrelated with each other and the two other inputs. They constitute the two cores in the organization and technology of the firm. If I_1, for example, changes, I_2 and I_3 are affected through feedbacks. Matrix (d) instead shows a perfectly modular production architecture in which inputs I_1 to I_4 are not related to each other, and inputs I_5 to I_7 are interfaces relating them. Improvements in any of the 'productive' inputs I_1 to I_4 do not influence any other, and the interfaces mediate only between two of the first four inputs. Thus the process technology's dependence on one or more activities is reduced; that is, the number of inputs with high K is reduced. This comes at the expense of more, yet highly specialized, inputs. Once such a design of the process technology has been achieved, the system can gradually be improved without incurring the risk of systemic collapse. It opens the way for a series of continuous improvements in the single inputs of the new set-up.

3. COMPLEMENTARITY CONSTRAINTS IN THE TRANSFORMATION OF OFFICE WORK, 1880–1930

The emergence of large administrative structures in manufacturing in the USA at the turn of the nineteenth century was the result of a process of adaption of firms to gradual and sustained changes in their economic environment. An information-processing administrative body within the firm became a necessity, as economies could be achieved through better coordination. The necessary change in entrepreneurial strategy led to a new structure of firms (Chandler 1962, 1977). Its operational implementation is discussed in Litterer (1963), Yates (1989), Levenstein (1998) and Hölzl and Reinstaller (2000).[2] We offer an interpretation of this process in accordance with the framework put forward in the preceding section.

3.1 The Development of a New Information System: the Accounting Revolution from the 1870s to the 1890s

The transformation of the enterprise from a purely productive entity towards the modern large enterprise with complex communication and coordination mechanisms took place gradually in response to a crisis of established management methods due to changes in demand, more complex production technology and problems of coordination on the shop floor. Cost control and accounting became a necessity. Cost accounting is a highly interdependent activity. Therefore the process to achieve an efficient cost-accounting information system took place in two overlapping stages. In the first the internal contracting system was replaced by bureaucratic business hierarchies. The second stage saw the transformation of the information-processing activities themselves as a response to constraints that were hit by the new information-processing system.

The typical office in the nineteenth century up to the late 1870s was virtually untouched by technology and consisted of predominantly male workers, a book-keeper, a copyist, a clerk and perhaps a shorthand taker. Clerical work had the characteristics of a craft where necessary skills were acquired on the job and through tutoring by seniors (Cooper and Taylor 2000, Braverman 1974). Accounting records reflected external market transactions and contained little information about internal operations. In larger firms, forms of inside contracting were the prevalent method of control. The contracting of internal craftsman avoided administrative overheads and acted partly 'as a substitute for accounting' (Hopper and Armstrong 1991, p. 415). A large office to process all types of cost and market data was not needed. The manufacturers set their prices on the basis of cost information but were not able to intervene directly in their determination, as they could not coordinate the production process.

The reaction to this failure was the systemic management movement that gained large support in US manufacturing in the late 1870s (Litterer 1963). It 'based its reassertion of control and co-ordination on record keeping and flows of written information up, down, and across the hierarchy' with the aim to 'transcend reliance on the individual in favour of dependence on system' and to monitor and evaluate performance (Yates 1989, pp. 10–11). Systemic management represented a set of procedures for decomposing activities into elementary work units.[3] The decomposition produced cost data, but also information of the exact relationship between work procedures and allowed introduction of standard measures of performance for single activities and sub-processes as well as their parcellation from or integration with others. The first step in taking control of the activities on the shop floor was the direct payment of wages and salaries. This led to the demise of the internal contracting system. The transfer of authority from foremen to plant managers was gradual. Larger production units supervised by a single foreman were broken up into smaller ones. Tasks such as the hiring and payment of workers, material acquisition and performance monitoring were centralized. Thereby the scope of authority of foremen was reduced to the allocation and monitoring of work within small production groups. Parcellation increased the number of foremen and work groups. Many decision tasks were integrated into new centralized staff departments, which acted as an interface between the shop floor and management. Output requirements were standardized, precise production schedules introduced and performance could be monitored through detailed cost figures. In this way management gained direct control (through integration) over activities on the shop floor. Accounting and its transformation into a current cost management technique were instrumental in this process. The importance of accounting was greatly increased, as it mediated the interactions between the different activities of production and distribution.

The formal bureaucratic structure was the result of a process of decomposition and recomposition with the hierarchy reflecting the information flow of accounting and market data, which in turn reflected the structure of the underlying production problems. The bureaucratic structure was a problem processor that allowed the decomposition of complex decision problems regarding the allocation of resources into sub-problems and at the same time led to a decentralization of decision-making and a centralization of ultimate control and coordination. But as the information system was set up, new constraints became binding for the information-processing activities themselves, which led, according to Beniger, to

> a crisis of control in office technology and bureaucracy in the 1880s, as the growing scope, complexity and speed of information processing . . . began to strain the manual handling system of large business enterprises. This crisis had begun to ease by the 1890s, owing to innovations not only in the processor itself (the bureaucratic structure) but also in its information creation or gathering (inputs), in its recording or storage (memory), in its formal rules and procedures (programming), and in its processing and communication (both internal and as outputs to its environment). (Beniger 1986, p. 390)

The large business administrations produced and processed information on an industrial scale. The economic survival of a firm depended on accounting and other administrative activities, which depended in turn on the trained clerks as necessary input. This kind of labour was in quasi-fixed supply. The soaring need for qualified clerks and the low potential for productivity advances made the constraints set by the labour market for clerical workers binding in the 1880s. In the three decades from 1880 to 1910 the share of clerical workers in the total working population increased from 1.1 per cent to 5.1 per cent: the number of clerks in the USA rose from 186000 to 1.8 million.[4] These changes in number went hand in hand with the transformation of clerical work and the reorganization of the office.

3.2 Finalizing the Information System: the Mechanization of the Administration from the late 1880s to 1930s

The sheer size of the information-processing volume and the labour-intensive character of clerical activity inflated the bureaucratic structure.

The systemic management movement itself had only marginally touched it, even though it was its product. On a smaller scale, these circumstances resembled the situation that had led to the development of the American System of Manufactures four decades before (Hounshell, 1984). Engineers played an important role in restructuring the bureaucratic machinery (for example McPherson 1992). Their primary objective was to reduce the dependence of administrative processes on skilled clerical labour by making labour and capital good inputs more separable. Their training in the tradition of American engineering provided the background for approaching this problem. The American System of Manufactures can be viewed as a meta-heuristic or problem-solving algorithm for problems in the production sphere. The model was the system of modular production based on standardized parts and activities. This model was applied to the organization. The division of labour in administrative work was increased through the standardization of tasks, data and information channels. In parallel, those activities for which this was possible were mechanized. The standardization of data and tasks was an important precondition for the introduction of office machines, as they could realize their full productivity potential only if a smooth flow of standardized and indexed information was available.

The role of parcellation and integration in this process becomes clear from the development of book-keeping practices. Book-keeping was divided into a sequence of distinct and specialized occupations. The change in methods, organization and processes took place gradually. The first step was to separate data handling, which was amenable to standardization, from data analysis, which was not. Through this the dichotomy of book-keeping and accounting emerged. Book-keeping activities were divided into activities of work preparation and activities of data manipulation. Work preparation tasks, such as the sorting of vouchers and receipts and the examination of related ledgers, were executed without mechanical aids, but reached a high degree of specialization in manpower. In most cases one clerk was responsible for some subset of tasks with respect to one narrow subclass of accounts. Data manipulation activities entailed the codification, processing and evaluation of data and were supported by mechanical equipment (Pirker 1962, pp. 81–3). Adding machines, calculators, book-keeping and billing machines or even Hollerith systems were adopted, which were operated by specially trained personnel. The organization of the accounting activities was turned into a modular system in which improvements or changes in one sub-activity were made independently from changes in other sub-activities. The increased division of labour reclassified clerical work into standardized and quasi-standardized activities.

The reorganization of administrative work and the invention and development of mechanical devices was a co-evolutionary process. 'It was not an accident of fate' (Cortada 1993, p. 63) that an office equipment industry was built up at about the time when large, multifunctional enterprises emerged by the early 1880s. Most information-processing devices were invented and introduced on the market between 1870 and 1890 (for example typewriters 1873, cash registers 1879, calculators 1885, Hollerith system 1889). They diffused about ten to 15 years later on a large scale. The key requirements of the new office machinery were summarized by Leffingwell and Robinson (1950, pp. 282–3): 'When should office machines be used? To save labour, to save time, to promote accuracy and to relieve monotony.'[5] One could add 'in order to achieve an order-of-magnitude jump in office work productivity'. The technical design of these office machines reflected the constraints out of which they were born, as they had to be instrumental in realizing a system of standardized activities. This is summarized in Table 6.1 for four distinct devices. The critical skills needed to perform quantitative or repetitive operations such as sorting or adding were embodied in some mechanism, and the machines could be operated without much previous training. This led to the saving of labour and increases in productivity, largely independent of the skills of the operators. Such an embodiment was not possible for more knowledge-intensive activities, such as typing and shorthand taking. In these cases office machines were used to support the specialization of labour. The result of this was a strong complementarity between the new technological artefact and the operator. But the standardization of the user interface such as the typewriter keyboard also forcibly led to the standardization of skills and the separability of the process from tacit knowledge or skills specific to a single worker.

Typewriters were the first technology of the new office work regime. Their domains of application were all activities involving the distribution of information on a small scale. The typewriter as a technical artefact was an innovation but did not in itself represent a productivity-increasing technical advance. Its mechanical construction did not embody any specific clerical knowledge or skills so that its use did not lead automatically to a productivity increase. The crucial criterion for the adoption of the typewriter was its (standardized) human–machine interface. The special interaction of service requirements and technological characteristics that gave birth to the QWERTY keyboard is well known (David 1985). The subsequent development of touch-typing played a crucial role in making the typewriter a viable technology for business administrations, as it contributed to the establishment of a homogeneous labour supply. The coexistence of different keyboards with different practices would have led to a segmentation of the

Table 6.1 Early information technology: innovation characteristics of the most important technologies of the IT regime

	User side			
Technology	(i) Supported economic function of organization and (ii) type of processed information	(i) Source of productivity gains and (ii) effect on established competences, i.e. clerical work before IT regime	New professions	Required skills
Typewriter	(i) Coordination (ii) Multiplying codifying of qualitative information	(i) User interface, touch typing and complementary technologies (ii) Replacement of copyists	Typist; establishment of typing pools	Touch-typing (about 60 words per minute), shorthand writing at least 60–75 words per minute (partly replaced by dictaphones), good language and grammar skills, letter-writing ability. High-school degree preferred. Training period: approx. 250–400 hours
Adding and calculating machines	(i) Monitoring (ii) Processing of quantitative (accounting) data	(i) Mechanical adding or calculating mechanism, automatic entry controls, user interface (ii) Replacement of mathematical skills	In large enterprises comptometer or adding machine operators; used in functionalized book-keeping, sales or billing departments also on sporadic basis. Establishment of computation pools	Machine use. Touch-typing. Training period: few days. Girls of about 17 years of age with two years in secondary school

Accounting machines	(i) Monitoring and allocation (ii) Processing of quantitative accounting data	(i) As for adding machines plus reduction of double-entry mistakes through better work preparation (ii) Replacement of book-keepers (mathematical skills, book-keeping skills)	None; used in functionalized bookkeeping departments which took also the form of bookkeeping pools	Machine use. Training period: accounting clerks with double-entry skills two weeks
Hollerith and Powers systems	(i) Monitoring and allocation (ii) Processing of quantitative (accounting) data	(i) Electric contact principle, codification of information, sorting and tabulating mechanisms (ii) Replacement of mathematical and statistical skills; sorting and indexing tasks	Card-puncher Sorter Tabulator Programmer Establishment of card punch units and machine rooms	Puncher: in some cases typing skills, mostly not; primary school degree. No further skills needed. Training period: 1–4 months Sorter: no special skills, but strong physical constitution required; primary school degree. Training period: about 6 months Tabulator: secondary school degree and technical skills. Training period: 1.5 to 2 years Programmer: organizational skills, business skills; preferably university degree in mathematics or a technical discipline. Training period: 4 years

Source: Hölzl and Reinstaller (2000).

labour market with an inevitably lower elasticity of supply. As typewriters were fixed capital, firms organized their operations in such a way as to maximize the rate of utilization. This led to centralized services for typing, and to functional office departments that pooled typing activities (Leffingwell and Robinson 1950, p. 34). Typing became a profession and an administrative process in its own right. The typist had a clearly defined activity profile, which consisted of taking (shorthand) notes and transferring them onto paper with the machine, using a particular typing method. Standards of practice were attained by the standardization of letter styles and forms, thereby influencing the way business correspondence was done (Leffingwell and Robinson 1950, p. 143). Typing pools were formed through the integration of new activities into the administration.

As the large administrative organizations owed their very existence to the attempt to gather and evaluate more quantitative data, the capability to perform simple mathematical operations was of foremost importance. Rationalization studies carried out at the time showed that almost 60 per cent of all tasks performed in office work consisted of calculations that amounted to 80 per cent of counting and adding (Pirker 1962, p. 66). Adding and calculating machines were general-purpose tools applicable to a vast range of uses. They embodied the most important skills of a good book-keeper: quick and reliable computing. The locus of the labour-saving potential was primarily the mechanical arithmetic unit, which was independent of the skills of the machine operator. The machines isolated mathematical operations completely from the persons carrying them out. Operators could learn quite quickly to handle them, but did not need to know much mathematics. They did not even need to know a specific standardized method such as touch-typing (see Table 6.1). As documented in Table 6.2, the labour-saving effect was in the order of 20 to 1. Long rows of numbers could be condensed into one single key figure much faster than before. Book-keeping machines entered the market in the 1920s. In most cases they were tailored to specific uses and were essentially combinations of an adding machine with a typewriter or just normal adding machines with mechanisms allowing special carriage movements.

Hollerith machines or tabulating gear differed from the previous technologies. Much more than adding and book-keeping machines, they were instrumental in the implementation of new organizational designs. This becomes clear if we consider that in order to sell such devices 'a salesmen had not to sell the machine but the organisation' (Pirker 1962, p. 79). Salesmen, who were typically trained engineers, acted as technical advisers as well as organization designers. The organizational concept developed for businesses in one particular sector was then used as a blueprint (and sale argument) for other firms in that sector (see McPherson 1992). This was

Table 6.2 Data-processing capabilities of early accounting machines c. 1926

Type of machine	Information-processing capacity,[a] bit/hour	Current account transactions per hour	Processing speed machine/clerk
Clerk, manual execution	270 (0.03 kbytes)	6	1
Burroughs adding machine	5400 (0.675 kbytes)	120	20
National cash register	6300 (0.78 kbytes)	140	23.3
Smith premier (typewriter combined with a calculator)	2475 (0.31 kbytes)	55	9.166
Elliot Fisher (typewriter combined with a calculator)	2475 (0.31 kbytes)	55	9.166
Ellis (typewriter combined with a calculator)	5400 (0.675 kbytes)	120	20
Underwood accounting machine	2925 (0.37 kbytes)	65	10.83
Hollerith or Powers card punch machine	6750–11250 (0. 84–1.4 kbytes)	150–250	25–41.6
Hollerith or Powers sorting machine	675 000–810 000 (84.375–101.250 kbytes)	15 000–18 000	2500–3000
Hollerith or Powers tabulating machine	162 000–202 500 (20.25–25.31 kbytes)	3600–4500	600–750

Sources and notes: [a] We used information contained in Meuthen (1926) for the calculations. We used the figure of a punched card for a current account transaction (p. 43) and the table of single cases that could be performed by the single machines (p. 48). The following information was stored on the punched card on 45 columns of 10 digits each for one transaction: date of the transaction (6 columns), the number and page of the main register (3 + 3 columns), the type of the transaction (2 columns), the department (2 columns), the main and secondary number of the current account (5 + 2 columns), debit and credit (8 + 8 columns) and finally the day and month of the booking (2 + 2 columns). The information in each of the 45 columns represents 1 bit (one of the ten rows is punched or not punched). This is only the information content of the card due to the punches and not the information content if the data were codified by binary numbers, which is likely to be much higher. We hence calculated the implicit information-processing capacity, and not the effective one. We further compare the transactions on the basis of how the most advanced technology (Hollerith: punched cards) of the time codified information. This entails that, with the purely manual system, information was not just written on one single card, but meant filling in fields in several different registers manually, for which a clerk on average needed 10 minutes for one single current account booking operation; ibid., p. 47.

necessary to utilize the capacity of this system at its operational optimum. Tabulating gear was used to process company-wide data on a large scale, shifting labour productivity for certain clerical tasks by three orders of magnitude, as documented in Table 6.2. Large business firms used these machines to tabulate sales statistics for payroll and inventory management, and later for consumer trend analyses that involved laborious counting and sorting processes.

A tabulating system consisted of punched cards (which were the media on which operating instructions and information were stored), cardpunch machines (which transferred the information on the cards), sorting machines and a tabulator to count the sorted cards. The sorters could be programmed with punched cards on which a sorting routine was codified. This made them very flexible, as they could be re-programmed. The operation of tabulating machines was split into three distinct activities: (i) the codification of sorting and tabulating routines; (ii) the codification of the information; and (iii) the evaluation itself, that is, the actual sorting and tabulating of information. Accounting and organization specialists carried out the codification of routines. Specific sorting and tabulating processes were stored on punched cards and could be used when necessary. The programming of routines and routine sequences was an activity that happened only sporadically at the set-up of the machine and subsequent organizational changes. These programs made codified procedural knowledge on clerical operations readily available. The codification and evaluation of information were, on the contrary, recurrent tasks.

Hollerith and Powers systems almost completely isolated productivity increases in data-processing from the skills of the manpower.[6] The high specialization and division of labour typical for the shop floor in the American System found its correspondence in a number of new occupations. Pirker (1962, p. 95) noted in regard to the organization of work that 'for the first time something appears in the office, that can be compared to the working practice on the shop floor'. Key-punch operators codified and controlled the information; sorters were responsible for the supervision of sorting and tabulating processes; lead-machine operators (also called tabulators) were responsible for the wiring of the control panels and the verification of the machines; and, finally, programmers were responsible for the cybernetic part of the job: programming and designing process flows. The skill requirements increased in ascending order: the skill requirements for key-punch operators and sorters were of primary school level, tabulators needed to have specific technical skills and therefore most had secondary school education. The few programmers needed were university graduates in mathematics or engineering, who were considered as professional organizers and operated in a middle management environment (see

Table 6.1). Unlike the other professions, the latter were in short supply, but their work was only needed sporadically.

The four technologies presented here were only a small part of a myriad of mechanical devices assuming a fixed place in the organization of office work, such as billing machines, cash registers, addressing machines, filing systems and so forth.

4. DISCUSSION

What emerges from the historical case study in the previous section is that there were two constraints that induced interdependent search processes for new technological and organizational designs. The first complementarity constraint was represented by the interdependence of the process generating cost information (accounting), and the process generating business decisions (management). The decentralized organization of production and the simple accounting system in use during the period of the American System of Manufactures was consistent with the management needs of the time. The system became inconsistent with a larger scale of operations. The change in management practice triggered changes in accounting practice. This is a clear indication of a complementary constraint. Once these organizations were set up, they produced information on a much larger scale, and the problem of efficient processing of the information emerged. The interdependence between production activities and information-processing and management needs led to a second constraint: more information could only be processed with more clerical workers, whose supply was quasi-fixed. This complementarity constraint was internal to information-processing activities and affected the functioning of the entire administrative machinery. As a result, information-processing activities were reorganized and new office machines developed and adopted.

The hypothesis now is that organizational innovations preceded innovations of technological artefacts. The inconsistencies between elements in the administrative process triggered the search for new solutions, a different method of information-processing and new technological artefacts in their support. An innovation count (Figure 6.3), which differentiates between organizational and technological innovations, lends some support to the hypothesis that organizational innovations led to technological innovations and that recombinant search was triggered and guided by complementarity constraints.

Figure 6.3 shows that in the field of information technology some clustering in inventive activity took place in the 1880s and 1890s and was preceded by a large number of innovations in different fields of organization.

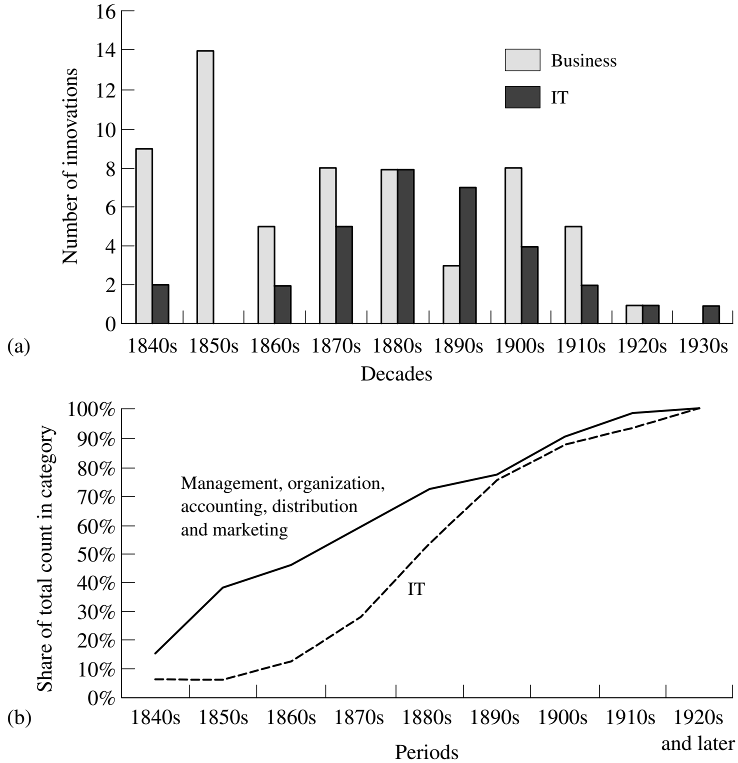


Figure 6.3 (a) Important innovations in business strategy and organization and in early information technology; (b) Cumulated innovations

Even if the data on which the graph is based were constructed carefully from a variety of sources, they may suffer some selection bias. These results should therefore be interpreted with caution.[7] Nevertheless, the data support the conjecture that technological search and innovation in the field of IT were induced by the complementarity constraints emerging from changes in business strategy. The clustering of innovations is the outcome of a directed search process.

The discussion of Section 3 showed that the design of the process technology of office work depended on the complementarity constraints that firms faced at the beginning of the process of adaption. The process of decomposition of tasks through parcellation and integration led to a more modular design of the administrative hierarchy. The new machines, which supported information processing, reduced dependence on skilled clerical work. The

Table 6.3 Economic effects of the first IT regime in US manufacturing: correlations

	Total period: 1889–1937	Subperiod
VA to A/P	0.685**	0.975*** (1899–1919)
VA to C_{office}/C_{total}	0.762**	0.858* (1899–1919)

Note:
*** Statistically significant at the 1% level, ** statistically significant at the 5% level, * statistically significant at the 10% level.

Source: Reinstaller and Hölzl (2001), provide a detailed account of data sources and definitions.

number of activities in the realm of the business administration grew and became more specialized, reflecting the more modular organizational design. In this phase a process of capital deepening and a rise in the capital stock used in the office took place.[8] The development and adoption of the first IT regime took place mostly in large firms, so that a significant difference in profitability measures between small and large firms is to be expected. And indeed, as Melman (1951) found for the period between 1899 and 1947, large firms were able to keep their administrative expenditures per dollar of production expense lower than small businesses despite the much more pronounced rise in the administrative overhead. Larger firms benefited more from their use than smaller ones, indicating that scale economies played an important role in the adoption decision. The results of a correlation analysis presented in Table 6.3 suggest that this hypothesis not only works on the firm level but also on a more aggregate level. The value added (VA) correlates positively with the share of clerks in total staff (A/P) and with the share of office machinery and furniture in total capital (C_{office}/C_{total}), suggesting that increases in value added were associated with the build-up of business administrations. The correlation coefficients are significantly different from zero and their values are close to one. This indicates that firms succeeded in their adaptive efforts to become more competitive through better cost control, which is implicit in the higher share of office workers in the total workforce. Simple correlation coefficients are of course no rigorous econometric test. These results should therefore be considered tentative. Better data should alleviate these reservations. Nevertheless, the results suggest that better coordination and organization of the shop floor increased the overall profitability of firms.

This chapter has attempted to trace out the nature of recombinant search by providing a framework of enquiry based on the NK model and by studying the historical case of the first IT regime. The case study and the descriptive statistical evidence taken together indicate that the search in the technology space was shaped by complementarity constraints and a response to the changing structure of economic incentives. These constraints were also reflected in the final design of the process technology of modern office work, as the search did not take place in the neighbourhood of the established technique. The observed patterns of technological development lend support to our theoretical conjectures. The results in this chapter are of a tentative nature. More detailed research is required to substantiate and generalize these findings.

NOTES

1. For a discussion see Fuss et al. (1978), pp. 244 ff.
2. For references and a more detailed account of the historical development see Hölzl and Reinstaller (2000), pp. 5ff.
3. Scientific management, which was introduced with moderate success towards the end of the 1890s, brought these principles to its limits; see Boorstin (1973, p. 369).
4. Compiled from J.M. Hooks' 'Woman's Occupations through seven decades', US Department of Labor 1947 and Historical Statistics, Abstracts of the US Series D57-71 Later: US Bureau of Census Statistical Abstract (1972). For details see Hölzl and Reinstaller (2000, p. 8).
5. These are the headings used by Leffingwell and Robinson (1950) to discuss the use of machines in office work.
6. Exceptions were key-punch operators. Their efficient operation relied on the speed with which the codification could take place. This first led either to the adoption of the typewriter keyboard (for alphanumerical insertions) or a ten-key keyboard (for purely numerical insertions) with the keys ordered in four rows. Both allowed for the use of touch-typing methods. As the codification of data continued to be the bottleneck in this technology, the use of standardized interfaces was important.
7. Information on the data is available upon request from the authors. Figure 6.3 shows innovations that have been classified as important by eminent economic historians. Organizational innovations have to be taken in the widest sense possible: they include new methods of production which induced change in the organization, such as the introduction of continuous processes, as well as innovations in the field of marketing, accounting and so forth.
8. By capital deepening we mean the use of more specialized and more productive capital goods.

REFERENCES

Altenberg, L. (1994) 'Evolving Better Representations through Selective Genome Growth', *Proceedings of the IEEE World Congress on Computational Intelligence*, pp. 182–7.

Altenberg, L. (1997), 'NK Landscapes', in T. Back, D. Fogel and Z. Michalewicz (eds), *Handbook of Evolutionary Computation*, Oxford: Oxford University Press.

Beniger, J.R. (1986), *The Control Revolution*, Cambridge, MA: Harvard University Press.

Boorstin, D.J. (1973), *The Americans: The Democratic Experience*, New York: Random Books.

Braverman, H. (1974), *Labor and Monopoly Capital. The Degradation of Work in the Twentieth Century*, New York: Monthly Review Press.

Caminati, M. (1999), *Complementarity, Opportunity and the Co-Evolution of Techniques*, Universitá di Siena, Quaderni Dipartimento di Economia Politica no. 260, July.

Chandler, A.D. (1962), *Strategy and Structure*, Cambridge MA: MIT Press.

Chandler, A.D. (1977), *The Visible Hand: The Managerial Revolution in American Business*, Cambridge, MA: Belknap Press of Harvard University Press.

Cooper, C. and Taylor, P. (2000), 'From Taylorism to Ms Taylor: The Transformation of the Accounting Craft', *Accounting, Organizations and Society*, **25**, 555–78.

Cortada, J.W. (1993), *Before the Computer: IBM, Burroughs, and Remington Rand and the Industry They Created, 1865–1956*, Princeton, NJ: Princeton University Press.

Cowan, R. and Foray, D. (1997), 'The Economics of Codification and the Diffusion of Knowledge', *Industrial and Corporate Change*, **6**, 595–622.

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

Frenken, K. (2001), 'Fitness Landscapes, Heuristics and Technological Paradigms: A Critique on Random Search Models in Evolutionary Economics', in D. Dubois (ed.), *Computing Anticipatory Systems*, Woodbury, NY: American Institute of Physics, pp. 558–65.

Frenken, K., Marengo, L. and Valente, M. (1999), 'Interdependencies, near-Decomposability and Adaption', in T. Brenner (ed.), *Computational Techniques for Modeling Learning in Economics*, Boston, MA: Kluwer, pp. 145–65.

Fuss, M., McFadden, D. and Mundlak, Y. (1978), 'A Survey of Functional Forms in the Economic Analysis of Production', in M. Fuss and D. McFadden (eds), *Production Economics: A Dual Approach to Theory and Applications*, Amsterdam: North-Holland.

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

Hölzl, W. and Reinstaller, A. (2000), 'The Adoption and Enforcement of a Technological Regime: The Case of the First IT Regime', Vienna University of Economics and Business Administration, Vienna, Working Papers Series Growth and Employment in Europe, No. 12, September.

Hopper, T. and Armstrong, P. (1991), 'Cost Accounting, Controlling, Labour and the Rise of Conglomerates', *Accounting, Organizations and Society*, **16**, 405–38.

Hounshell, D. (1984), *From the American System to Mass Production*, Baltimore, MD: Johns Hopkins University Press.

Kauffman, S.A. (1993), *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford and New York: Oxford University Press.

Leffingwell, W.H. and Robinson, E.M. (1950), *Textbook of Office Management*, New York: McGraw-Hill.

Levenstein, M. (1998), *Accounting for Growth: Information Systems and the Creation of the Large Corporation*, Stanford, CA: Stanford University Press.

Litterer, J.A. (1963), 'Systematic Management: Design for Organizational Recoupling in American Manufacturing Firms', *Business History Review*, **37**, 369–91.

McPherson, J. (1992), 'Computer Engineer, an Oral History Conducted in 1992 by William Aspray', IEEE History Center, Rutgers University, New Brunswick, NJ, 29 April and 12 May.

Melman, S. (1951), 'The Rise of Administrative Overhead in the Manufacturing Industries of the United States', *Oxford Economic Papers*, **3**, 62–112.

Meuthen, J. (1926), *Die Maschinenarbeit in den Deutschen Banken*, Berlin: Julius Springer.

Pirker, T. (1962), *Büro und Maschine*, Basel: Kyklos Verlag.

Reinstaller, A. and Hölzl, W. (2001), 'The Technological Bias in the Establishment of a Technological Regime', Maastricht: Maastricht University, MERIT-Infonomics Research Memorandum 2001–014.

Saviotti, P. and Metcalfe, J.S. (1984), 'A Theoretical Approach to the Construction of Technological Output Indicators', *Research Policy*, **13**, 141–51.

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

Schumpeter, J.A. (1947), 'The Creative Response in Economic History', *The Journal of Economic History*, **7**, 149–59.

Simon, H.A. (1996), 'The Architecture of Complexity: Hierarchical Systems', in *The Sciences of the Artificial*, Cambridge, MA: MIT Press.

Tushman, M.L. and Anderson, P. (1986), 'Technological Discontinuities and Organizational Environments', *Administrative Science Quarterly*, **31**, 439–65.

Von Hippel, E. (1990), 'Task Partitioning: An Innovation Process Variable', *Research Policy*, **19**, 407–18.

Wagner, G.P. and Altenberg, L. (1996), 'Complex Adaptions and the Evolution of Evolvability', *Evolution*, **50**, 967–76.

Yates, J. (1989), *Control through Communication: The Rise of System in American Management*, Baltimore, MD: Johns Hopkins University Press.

7. An evolutionary approach to the theory of entrepreneurship

Thomas Grebel, Andreas Pyka and Horst Hanusch

1. INTRODUCTION

Research on entrepreneurship has always been a controversial topic in economic theorizing. The significance of entrepreneurship is emphasized by almost all authors working on innovation economics; nevertheless, most of the research work comes to an end at a purely appreciative level. Still, a consistent theory of entrepreneurship is missing – a theory that is adequate to combine the various strands of literature in order to arrive at an empirically testable model, eventually. Apart from the early theories that approach entrepreneurship from a rather intuitive perspective, tracing back to Schumpeter (1911, 1939) and Kirzner (1973, 1999), a modern evolutionary approach should also contain some specific theories such as the theory of human capital (for example Schlutz 1975, social networks, for example Granovetter 1973) and neo-Schumpeterian economics (for example Loasby 1999). In this chapter we develop an eclectic approach by designing an analytical model that can be applied to different industries and historical settings.

The core element of our model is the actors. Even though there are two opposing views that either explicitly focus on actors or take a more general approach in only emphasizing the actors' environment, for our purposes we draw on the actor-centered perspective. Therefore, we do not look at actors from the perspective of situative determinism and optimal behavior, but we characterize individual actors as procedural rational, struggling in a trial-and-error process for survival and prosperity. Consequently, in their entrepreneurial decision they do not know the potential economic outcome but rather experimentally try different combinations. The actors in our model are heterogeneous in their individual endowment concerning their accumulated human capital, their available venture capital, and their entrepreneurial attitude. We furthermore stress the importance of individuals' networks in the process of firm foundation as a social phenomenon. The formation of social networks is approximated by a random permutation process within our population of actors. In detail, an arbitrary number of actors, not yet involved in a firm, are randomly matched in each period. The comprehensive endowment of the group's actors constitutes their potential to found a firm. Whether a new firm is founded or not depends on their respective environment. In particular, they take into account the industry's economic development. As they obviously do not have perfect knowledge of all critical factors which drive an industry's development, they evaluate the average industry's performance by a chosen set of economic indicators. These evaluation criteria can be seen as the actors' threshold to establish a firm. Only in those cases where the actors' perceived comprehensive endowment appears to be sufficient to enter a market and the expected economic future signals promising rewards is a new firm born. When this happens, the birth process has an influence on the industry level, which in turn has a feedback effect on the micro-level, which is the effect on other agents' future decisions. Thus we manage to model a micro–macro relationship essential for the endogenous evolution of the foundation threshold, which takes place in historical time. Whereas the act of founding a firm depends on the individuals' perceptions and on the evaluation of their current (micro- and macroeconomic) situation, the firm's economic success, once founded, is determined by the individuals' resources and their specific managerial capabilities, which are embedded in the combination and complementarities of their skills – in short, their human capital. Accordingly, in the short run the survival of the firm decisively depends on a balanced relationship between human capital and venture capital. Missing human capital cannot be substituted by venture capital and might eventually lead to insolvency. As the firm has to invest its funds profitably within a certain period of time, it faces a bottleneck in human capital in the case of a maladjusted relationship. In the long run, however, the economic success of a firm depends on its competitiveness, which in our model – for the sake of simplicity – is determined by its stock of human capital and its learning capabilities to improve on it. If the firm has to exit, this again has a feedback effect on the foundation threshold. Due to the heterogeneous composition of actors and their experimentally organized behavior, our model has been illustrated numerically for the time being. In our first simulation experiments we are able to show the emergence of new industries and their endogenous evolution from a theoretical stance. Firms do not appear continuously but in swarms, showing a high degree of sensitivity to the coincidence of entrepreneurial behavior and environmental conditions. Our model is designed in a very general way and the promising results achieved so far advocate applying this basic setting to recent empirical observations of developments of new industries. Finally, this should improve our knowledge of conditions favoring/hindering the development of successful knowledge-intensive industries such as the information technology and biotechnology sector, respectively. This will be the agenda for our future research.

2. THEORETICAL MOTIVATION

2.1 A Historical Sketch and a Conglomeration of Entrepreneurial Functions and Ideas

The importance of entrepreneurial behavior for economic development has always been stressed in economic history, but the existence of entrepreneurship in orthodox economic theory has been almost undetectable. Economists wonder why the entrepreneur has almost vanished in economic theory.[1] The apparent reason is that, with the introduction of entrepreneurial behavior in orthodox theory, a model runs the risk of losing its consistency, and therefore the entrepreneur remains a stranger in economic theory. Classical economists have touched on this subject matter more than neoclassical theory, which is based on the equilibrium concept, might ever be able to do. Its strict methodological apparatus appears to rule out the possibility of picking out an endogenous equilibrium-disturbing element as the centerpiece of economic development.

The first to consider the role of entrepreneurs in the economy was Cantillon (1680s–1734).[2] He classified economic agents into three groups: (1) landowners, (2) entrepreneurs and (3) hirelings.[3] Whereas the first and the third groups are characterized as rather passive, the entrepreneurs play the central part in his *Essai sur la nature du commerce en général*. They play the role of the coordinator connecting producers with consumers and, additionally, the role of the decision-maker struggling with uncertainty and engaging in markets to earn profits. Cantillon's concept of uncertainty was however constrained to the entrepreneurs and the field had to wait for Frank Knight (1921)[4] for a detailed distinction between risk and uncertainty as an economy-wide feature affecting all economic agents. Cantillon was also the first to emphasize the entrepreneur's economic function while distinguishing it from the agent's social status. A functional perspective was maintained by those of Cantillon's successors associated with the French school. Quesnay,[5] the precursor of the 'Physiocrats', shifted the field of concentration to the significance of capital for economic growth. He thereby reduced the role of the entrepreneur as opposed to the industry leader to that of a purely independent owner of a business, even if he was endowed with individual energy and intelligence.[6]

Baudeau (1919)[7] was the first to propose the function of the entrepreneur as an innovator and thus brought invention and innovation into the discussion. Furthermore, he emphasized the ability to process knowledge and information, which makes the entrepreneur a lively and active economic agent. Another rather capitalistic view was propounded by Jacques Turgot[8] (1727–81). According to him, the entrepreneur is the outcome of a capitalist investment decision: the owner of capital can either simply lend his money and just be a capitalist, or decide to buy land for lease and hence become a landowner, or he can decide to buy goods to run a business and thus automatically become an entrepreneur. Jean-Baptiste Say (1767–1832)[9] developed Turgot's ideas and elevated the entrepreneur to a key figure in economic life. In contrast to Turgot he made a sharp distinction between the entrepreneur and the capitalist. The entrepreneur might give capital to a firm but he does not have to. Consequently, this also allows him to ignore risk and uncertainty,[10] when explicitly considering the entrepreneurial element. Say suggested a twofold approach. He looked at the entrepreneur from an empirical perspective to find out the actual entrepreneurial behavior – this he tried to reduce to a general entrepreneurial theory in a second step by subtracting all incidental aspects attributable to certain social and institutional circumstances.[11] The function of his entrepreneur was to understand technology and to be able to transfer that knowledge into a tradable product that meets customers' needs.

Say paved the road to Schumpeter's theory on entrepreneurship. And Schumpeter's entrepreneurial concept has to be seen as the pivotal point in this field of research. With a few exceptions, most of the economists before Schumpeter worked within equilibrium theory and most of the theories on entrepreneurship after Schumpeter are built on his ideas.[12]

Before we proceed to the discussion of Schumpeter's concept, we must briefly review the neoclassical treatment of the entrepreneur.

2.2 Neoclassical Constraints

The question 'What is the place of the entrepreneur in orthodox neoclassical theory?' is easy to answer, but it takes quite an effort to lay out the argumentation. The answer is: there is no space for an entrepreneur in neoclassical theory. The relevant discussion can be found in Barreto's work *The Entrepreneur in Microeconomic Theory* (Barreto 1989), where he portrays the disappearance of the entrepreneur in economic theory.[13] He shows that with the advent of the modern theory of the firm, economists lost track of the entrepreneur. Basically, the assumptional framework does not allow for a consistent implementation of entrepreneurial behavior. The bone of contention is rooted in the perfect rationality assumption which is a necessary condition for optimal behavior. This does not allow for a 'real' choice and the treatment of true uncertainty subject to entrepreneurial behavior, which limits the role of the entrepreneur to a static and passive and therefore redundant economic agent within a self-running firm. It is beyond the scope of this chapter to recount the anamnesis of the entrepreneur in economic theory. Basically, Schumpeter's legacy can be regarded as the outcome of such reflection.

2.3 Schumpeter's Entrepreneur

To give an accurate account we must go back earlier than Schumpeter's concept of the entrepreneur. As mentioned above, Schumpeter's work was tremendously influenced by a critical review of equilibrium theory. Though Schumpeter was fascinated by Walras's system of equilibrium, he stated that equilibrium theory contributed as much as it can; but further insights cannot be expected.[14] Schumpeter's circular flow is a less formal representation of Walras's general equilibrium theory.[15] To reach equilibrium, Schumpeter suggests that economic actors' decisions and actions have to be repeated over and over again in the same way, so that eventually all actors' plans coincide to end up in equilibrium. Schumpeter calls this result a static situation that does not allow for change.[16] His aim was to investigate the dynamics behind empirically observable economic change. He called the explanatory element *innovations*, and he called the economic agent who would bring along *innovations* (that is, 'new combinations') the entrepreneur.

When we look back at the existing literature at that time, Schumpeter's entrepreneurial concept is a synthesis of, first, Say's and Badeau's work and, second, the critique associated with the Austrian School.[17] Schumpeter's entrepreneur was and still is the most renowned concept. Therefore, we also take it as the intellectual foundation of our model. Another economist who should be mentioned in this context is Israel Kirzner.

2.4 Kirzner and the Austrian School

There is a long-lasting debate, partly stirred up by Kirzner himself, about the significant difference between Schumpeter's and Kirzner's entrepreneur. Both Schumpeter and Kirzner took up the Austrian critique of general equilibrium theory. Whereas Schumpeter developed a, to our minds, more general approach to entrepreneurship starting with economic change, Kirzner focused on the market process. For the reader's convenience, the intuition of the Austrian school is briefly recalled: equilibrium theory neglects market processes. If all plans of economic actors match, then there is no need for markets. In a state of disequilibrium, however, actors' plans do not match. They have to be revised and adapted to the new market situation.[18] The economic agents have to change their minds continuously and thus generate a dynamic process which Kirzner calls the *market process*.[19] This suggests that carrying out a Robbins type of maximization calculation[20] is impossible. Von Mises[21] solved this task by introducing *human action*.[22] Aside from their attempts to solve economic problems, agents are also alert to opportunities. Once an economic agent recognizes a market opportunity, he acts on it to improve his position. And opportunities are abundant in a situation of disequilibrium. That is where Kirzner's entrepreneur originates. While von Mises admitted the ability of human action to every economic agent, Kirzner confined it to a certain group of agents which he labeled entrepreneurs. Hence the concept of the entrepreneur as an arbitrageur who equilibrates markets was born.[23]

2.5 A Word on the Schumpeter–Kirzner Entrepreneur Discussion

Kirzner himself[24] distinguished the Schumpeterian entrepreneur as the innovator and the creative destroyer of equilibrium, from his own concept of the equilibrating entrepreneur alert to market opportunities. We leave it to the reader to distinguish between an equilibrium-disturbing and an equilibrium-creating entrepreneur. As a matter of perspective, if we see the alertness to market opportunities and the agent's implied human action as a part of innovativeness, neglecting the question whether a state of equilibrium in a dynamic economic world will ever be reached before another dynamic entrepreneur comes to prevent the economy from equilibrium, it would leave us with the centerpiece of the Schumpeterian dynamics of economic change, that is, the entrepreneur. In short, Schumpeter's stream of consciousness is as follows: no entrepreneur – no innovation – no dynamics – no evolution.

2.6 'Giving up the Grail'

Ever since economists started theorizing on human behavior, they have been looking for consistency in theory. What classical theorists could not achieve, neoclassical economists succeeded in. The marginal school and in particular Walrasian general equilibrium theory eliminated the shortcomings in terms of inconsistency within economic theory. They managed to refine the patchwork of classical thoughts to a consistent unity, but, as we have seen in the discussion above, at the cost of the entrepreneur. On the contrary, if we give up on the equilibrium concept for the sake of the entrepreneur, we might run the risk of losing consistency in return. Then we must practice disequilibrium economics without the powerful mathematical apparatus of the neoclassical school. *Equilibrium* needs *optimal behavior*. *Optimal behavior* needs *perfect rationality*. *Perfect rationality* requires *perfect foresight* and *information*. Regardless which of these assumptions we relax, we at the same time question the validity of the remaining ones, and, even worse, we question the methodological approach. This all foreshadows another era of patchwork in economic theory in what concerns the investigation of entrepreneurship, until an appropriate methodology is found. These misgivings are confirmed when we look at the existing literature which refers to entrepreneurship and at the same time abandons the equilibrium concept.

3. AN EVOLUTIONARY APPROACH TO AN EVOLUTIONARY CONCEPT OF THE ENTREPRENEUR

> the word 'evolutionary' is extremely vague. It is now widely used, even by economists using neoclassical techniques. 'Evolutionary game theory' is highly fashionable. Even Walras is described as an evolutionary economist (Jolink, 1996). . . . In precise terms it signifies little or nothing.[25]

For this reason, we decided to discuss briefly what 'evolutionary' means to our minds. The model presented in this chapter is intended to be a general approach to entrepreneurship, delivering constructive propositions for a basic evolutionary setting.

Consolidating the critique of Schumpeter, the body of thought from the Austrian School and, accordingly, Kirzner's adaptations to the entrepreneurial case, research on entrepreneurship becomes the pivotal point of a micro-based evolutionary theory. Many issues are addressed that boil down to questioning the phenomenon of innovation in an economic system. Innovation means novelty and, in accordance with Arrow's epistemological reservation, an assumptional fragile framework built on perfect foresight (complete information), which means perfect rationality, is a contradiction in itself. It ignores economic change spurred by the dynamic entrepreneur. Each of the assumptions mentioned above entails a huge discussion leading to various strands of literature. Of course, it is not the aim of this chapter to cover all of it, but they have to be taken into account, implicitly.[26]

Along these lines, we begin at the micro-level in our model. The agents are heterogeneous and differ in their individual endowment. Information is incomplete, in particular with respect to future economic development.

Because of imperfect foresight, agents have to deal with true uncertainty.[27] Furthermore, the bounded rational[28] agents are limited in their cognitive capabilities to perceive and process the accumulated information. Owing to the high degree of novelty attached to entrepreneurial behavior, true uncertainty does not allow for a calculation of expected values. The agent neither knows the set of possible outcomes nor the corresponding probabilities. As we thus deprive the agents of optimizing capabilities, they have to make decisions to the best of their knowledge. They have to perform, in the words of von Mises, *human action*. The agents, therefore, have to form expectations on various respects: first, they have to evaluate their individual endowment of resources, capabilities and competencies; second, the possibilities to acquire missing complementarities (to be specified later on); and third, the 'economic situation'.

Without the light of perfect rationality, agents consequently make individual forecasts motivated by their personality[29] and current (economic) environmental factors. In other words, decisions are the outcome of a path-dependent process: the evolution of the agent's individual (accumulated) endowment (resources, capabilities, competencies (including experience)) and non-individual, environmental factors subsuming the economic situation. The latter gives us the notion of feedback effects. The economic agents' decisions are influenced by economic factors (economic situation) and in return influence economic factors by their actions, that is, by the decision to establish a firm. It goes without saying that we implicitly consider irreversibility to round off the assumptional frame of our evolutionary perspective.

In the following, we flesh out this view with some less abstract ideas of entrepreneurial behavior. Since our main intention is to show the basic structure of an evolutionary model of entrepreneurship, we decided to tolerate some simplifications.

3.1 Actors

We divide an agent's individual endowment into three components which we call entrepreneurial spirit, human capital and venture capital. These three factors form the individual agent.[30]

3.1.1 Entrepreneurial component

The entrepreneurial component can be thought of as the residual of the agent's individual endowment which withdraws itself from empirical measurability. It comprises the intangible characteristics of the heroic Schumpeter entrepreneur. In this, we follow empirical evidence that does not allow us to detect a stereotypical entrepreneur and we furthermore accept von Mises's saying that every human being has the potential of human action.

3.1.2 Human capital

The second component refers to one of the more successful strands of research. The human capital approach, constituted by Theodor W. Schultz[31] and elaborated by Gary S. Becker, among others,[32] allows for an empirical application. It tries to explain optimal investment in human capital and delivers insights on income distribution. The theoretical concept is basically derived from investment theory in physical capital using marginal analysis. We do not use the human capital concept this way,[33] but we emphasize the importance of human capital for establishing a firm. Agents do not know the actual return when they decide in favor of founding a firm; they might do so when offering their human capital to the labor market. Therefore, agents decide in a dichotomous way: if they expect the returns of becoming an entrepreneur will be higher than those of being an employee, they will decide to become an entrepreneur.

Moreover, we refer to the literature on knowledge originating from the Austrian School,[34] which discusses the importance of knowledge in a disequilibrium situation – that is, a situation of uncertainty. Loasby (1999) gives a good intuition in *Knowledge, Institutions and Evolution in Economics*.[35] For our purposes, we interpret the agents' role of human capital as the crucial productive element for the long-run survival of the firm, once it is founded by the agents.

3.1.3 Venture capital

The third element we include in the agents' endowment vector is that of venture capital. We herewith pay tribute to the discussion on whether the roles of capitalist and entrepreneur can be separated. While the 'early French view' saw the entrepreneur as a risk bearer, the 'English view' identified the entrepreneur as a capitalist. Schumpeter (1939) discusses the role of money, too.[36] The bottom line is that potential entrepreneurs need to have capital to start their business, no matter whether they own it themselves or borrow it from others. Empirical evidence supports the hypothesis that entrepreneurs in general face financial and liquidity constraints.[37]

The intuition we draw from this discussion is that we assume each agent to be endowed with a certain amount of capital he can spend on a business venture. Again, we do not bother about the details, for example whether he inherited or accumulated a certain amount of money by saving.

So far, we have characterized individual agents by their endowment factors.[38] Each actor possesses the potential to be an entrepreneur, as von

Mises suggests from a theoretic perspective and, as empirical data show, most agents have.[39] Thereby the decision (human action) is no behavior of optimality, calculating what the maximal return to total – human and (free disposable) venture capital is. Yet the long-run survival of a firm, once founded, is highly dependent on the agent's human capital. As we vested all agents with the option to own venture capital, we can incorporate the notion of risk bearing and uncertainty. But as we will see later on, the 'mainly'[40] entrepreneurial agent need not be the risk bearer.[41]

Even though an agent might have a certain amount of every component necessary to establish a firm, he might not have enough of it. In that case, the agent needs to complete the minimum endowment he thinks necessary. Consequently, we introduce a network approach to entrepreneurship.

3.2 Social Networks

To complete the minimal endowment – although it still has to be defined what minimal means – actors can choose several ways to acquire such factors. They have to figure out how to get access to required resources (Penrose 1959) and whether the necessary competence to combine these resources (Foss 1993) is available. To draw on Coase (1988), some of the resources and competencies can be inherent in the agent; others have to be acquired in the market or otherwise. We will not go further down this road and we leave that task to a modern evolutionary theory of the firm still to be developed,[42] as we do not argue on the firm level but, following Birley (1985), investigate the 'pre-organization' phase in order to stress the importance of an agent's social network as a main source of help in obtaining resources and competencies to start a business.

Furthermore, we discuss the role of social networks[43] for two reasons. The first is methodological: by implementing social networks in the model, we 'climb up the aggregation ladder' one step further and, thus, leave the micro-level (individual's level) to bring the agent's social context into the discussion. The second reason is empirical: personality-based theories, that is, purely micro-based theories, try to find personal traits unique to entrepreneurs.[44] Nevertheless, these attempts have not yet been successful in identifying the entrepreneur when not considering the social group context.[45]

It is beyond the scope of this chapter and not our intention to comprehensively discuss social network theory. As we put together existing fragmentary theories on entrepreneurship in an evolutionary model setting, we incorporate social networks as another critical element in entrepreneurial behavior.[46]

4. THE MODEL

In the following section we introduce the basic structure of our model of entrepreneurship evolution. The model is designed in a very general form so that it will eventually allow us to investigate different scenarios and furthermore to implement the relationships and specificities of certain sectors. In a way the basic design has to be seen as a platform approach allowing several extensions with regard to the theoretical perspective as well as with regard to a close look at the empirical sphere.

4.1 The Actors

To model the evolution of entrepreneurship and the founding of new firms, we obviously have to go one step further down towards the micro-level, that is, not only down to the firm level but to the individual actors' level and in particular to the individuals' specific endowment. The individuals are characterized by the crucial features identified in the previous section: (i) entrepreneurial spirit es_i^t, which describes an actor's tendency not to become an employee but an independent firm leader; (ii) human capital hc_i^t, representing an actor's specific level of technological as well as economic knowledge and skills; and finally (iii) the actor's endowment and/or access to venture capital vc_i^t. These different features are all represented as real numbers on a cardinal scale in the interval [0,1], with higher values indicating higher levels of the specific characteristics. Accordingly, the n different actors in our model are described by the following triple:

$$a_i^t = \begin{pmatrix} es_i^t \\ hc_i^t \\ vc_i^t \end{pmatrix}, \tag{7.1}$$

where a_i^t: = actor i at time t, $i \varepsilon \{1, \ldots, n\}$. To build a starting distribution of the population of actors (Equation 7.2), we randomly create n of these triples where the features es_i^t, hc_i^t and vc_i^t are uniformly distributed within the relevant interval:

$$A^t = \{a_i^t\}_{i \in \{1, \ldots, n\}} \tag{7.2}$$

4.2 MATCHING PROCESS AND FOUNDING THRESHOLD

For each iteration, the population of actors not yet involved in a firm is permuted and k different actors are randomly brought together in order to evaluate their chances of founding a possibly successful firm. For this purpose, we consider the specific attributes of the actors to be additive, so that a potential firm pf_q^t can also be characterized by the triple of attributes of its k members:

$$pf_q^t = \begin{pmatrix} \sum_{i=1}^{k} es_i^t \varepsilon k_q^t \\ \sum_{i=1}^{k} es_i^t \varepsilon k_q^t \\ \sum_{i=1}^{k} es_i^t \varepsilon k_q^t \end{pmatrix}, \tag{7.3}$$

so that the set of potential firms at time t is

$$PF^t = \{pf_q^t\,(=ce_q^t)\}_{q\in\{1,\ldots,m\}}, \tag{7.4}$$

where $q \varepsilon \{1, \ldots, m\}$ denotes the specific potential firm and m the number of potential firms, that is, the number of temporarily formed k-groups q in period t. Each group of actors has to evaluate whether their comprehensive endowment ce_q^t, which for the sake of simplicity is equal to pf_q^t, is adequate. Yet the actors' mere perception of their common resources, attitudes and motivation is not the only determinant for founding a firm. The actors involved are also influenced by their environment and the respective mood within the population. For modelling reasons, we introduce the so-called founding or entry threshold Ψ^t, a 'meso-macroeconomic signal' which endogenously depends negatively on the growth rate of the sector's turnover w^t which in return reduces the threshold. Furthermore it depends positively on the rate of exits d^t of firms in the respective industry which increases the threshold:

$$\Psi^t = \Psi\left(\frac{dw^t}{dt}, d^t\right). \tag{7.5}$$

If the k-group's, that is the potential firm pf_q^t's, comprehensive endowment ce_q^t exceeds the foundation threshold Ψ^t, the k actors decide to found a firm; thus the potential firm pf_q^t turns into an actual firm f_j^t, and the formerly potential firm's comprehensive endowment ce_q^t becomes the actual founded firm's comprehensive endowment ce_j^t. Equation (7.6) gives the set of newly founded firms F_{new}^t in period t:

$$F_{new}^t = \{pf_q^t : pf_q^t > \Psi^t\} pf_q^t \varepsilon PF^t \tag{7.6}$$

Hence the set of all firms that have been founded up to time t is given in Equations (7.7), and (7.8) gives the firm j's comprehensive endowment.

$$F^t = \{f_j^t\}_{j\in\{1,\ldots,xt\}} \Leftrightarrow \bigcup_0^T F_{new}^t \tag{7.7}$$

$$f_j^t = ce_j^t = ce\left(\sum_{i=1}^{k} es_i^t, \sum_{i=1}^{k} hc_i^t, \sum_{i=1}^{k} vc_i^t\right)_{j \in \{1,\ldots,xt\} i \in a} \tag{7.8}$$

If the threshold is not exceeded, the option to found a firm, for the time given, is rejected by the actors. Consequently, the actors that do not get engaged in a firm are free to go for further trials in the following period. In the case of a successful foundation of a firm f_j^t with $j \varepsilon \{1, \ldots, x^t\}$ the k actors involved are no longer available to found another firm. At the same time, this reduces the probability for other actors to find adequate partners. On the other hand, according to Equation (7.9) the number of existing firms x^t is increased by the number of firms F_{new}^t founded within a period, thereby also exerting a positive influence on the sector's aggregate turnover which positively feeds back on the founding threshold in the next period.

$$x^t = x^{t-1} + |F_{new}^t|, \tag{7.9}$$

where x^t: = number of firms in the industry at time t.

4.3 Survival and Exit

Whether a firm f_j^t survives in the market or is threatened by exit critically depends on its set and composition of aggregated capabilities as well as on the turnover that firm is able to acquire. Most simply, we assume for a firm's turnover that it is positively influenced by its aggregated human capital $hc_j^t = \Sigma_{i=1}^{\infty} hc_j^t$ and its aggregated venture capital $vc_j^t = \Sigma_{i=1}^{\infty} vc_i^t$. In order to approximate a positive influence of experience the accumulated turnovers $\Sigma_t{}_j^{t-1}$ also exert a positive influence:

$$w_j^t = w\left(\alpha \cdot hc_j^t, \beta \cdot vc_j^t, \gamma \cdot \sum_t w_j^{t-1}\right), \tag{7.10}$$

where α, β, γ: = weighting parameters.

Finally, in cases where the composition of a firm's f_j^t specific characteristics shows an unfavorable relation with respect to a relatively low amount of human capital hc_j^t compared to its venture capital vc_j^t, we introduce a so-called burning rate w_j^t (Equation 7.11) which depreciates the available capital over the course of time according to Equation (7.12):

$$w_j^t = \begin{cases} 0 & \text{if } \frac{hc_j^t}{vc_j^t} \geq 1 \\ \frac{hc_j^t}{vc_j^t} & \text{if } \frac{hc_j^t}{vc_j^t} < 1 \end{cases} \tag{7.11}$$

$$vc_j^t = vc_j^t - \sigma \cdot w_j^t, \tag{7.12}$$

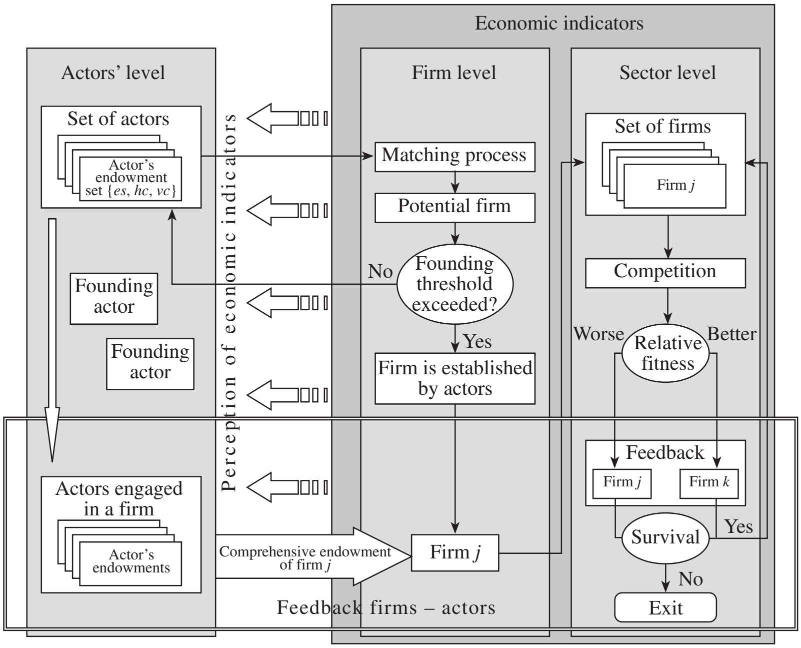


Figure 7.1 Basic structure of the model

where σ: = weighting parameter. A firm has to exit the market in the cases when its venture capital vc_j^t is reduced to zero.

4.4 Basic Structure of the Model

Figure 7.1 summarizes the basic structure of the model.

To start, we distinguish several levels of analysis: the actors' level, the firm level and the sector level. The entrepreneurial process primarily takes place on the actors' level. A set of actors with heterogeneous endowments is given. Actors form social networks that change over time, expressed by a random matching process.

The actors, grouped together, constitute a potential firm. Since they neither have perfect foresight nor complete information about future prospects, their decision will be myopic and based on their common evaluation of the economic situation, which is influenced by their subjective perception of measurable economic indicators. The more economic indicators paint a promising picture of a possibly prosperous outcome of entrepreneurial actions, the lower the threshold for actors to decide in favor of such action. The same holds vice versa. If actors decide against founding a firm, they return to the set of actors available for another trial to evaluate entrepreneurial actions within a changed social environment. If they decide to found a firm, the firm is established and actors' resources are bounded within the firm so that they are excluded from a further firm-founding process. On the sector level, the firm is forced to compete with incumbent firms. Their competitiveness is determined by their comprehensive set of endowments, constituted by the founding actors' individual endowments. The selection process, which is competition, has an effect on each firm, either worsening or improving its fitness to stand future competition. The short-run exit criterion, competing 'for the market', is insolvency. Firms with an unbalanced set of endowments run out of money and finally have to exit the market. The long-run selection process via market competition, or in other words 'competition in the market', decides the competitiveness of the actual business idea.

5. RESULTS

In this section we present some of the first simulation results of the model. Although our focus is on entrepreneurial behavior, we have to take a rather holistic view. Combining the manifold theoretical contributions in the realm of the analysis of entrepreneurial behavior, we also have to touch on some peripherals of the subject investigated in order to show the endogenous dynamics of entrepreneurship. Otherwise, it would not be possible to include the feedback effects suggested in the model. Nevertheless, we neglect a further specification of those peripheral, economic phenomena and leave it with a purely theoretical case. The simulations we ran all show the same qualitative features. A Monte-Carlo simulation to support our findings still has to be carried out. To start with a stereotypical development of the emerging sectors' total turnover, see Figure 7.2. Once firms are founded, the industry's total turnover increases sharply. The high growth rates at the beginning function as a signal for other economic actors to enter the market (to innovate), too. From a certain point in time, as competitive pressure increases, with more and more firms entering the market and as market diffusion proceeds, growth rates decline while remaining positive. Thus the total turnover curve takes a stylized sigmoid shape. Firms do not enter all at once. Some enter early whereas others enter later. Early entrants might have a first-mover advantage, whereas late entrants might have to struggle for survival, competing with larger firms. It is not just the time of entry that makes firms different but also their set of endowments

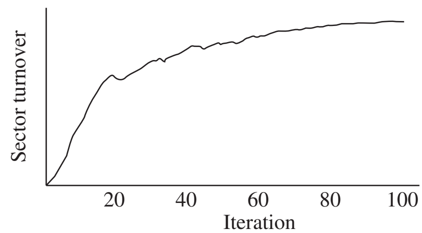


Figure 7.2 *Overall turnover of the emerging sector*

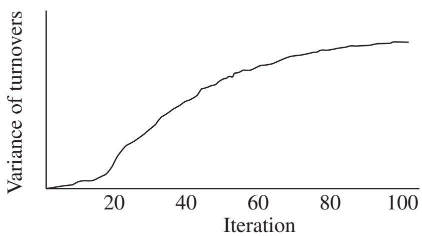


Figure 7.3 *Variance in turnover*

that is crucial for their economic development. Figure 7.3 shows the heterogeneity of firms by indicating the emerging sectors' variance in turnover.

Taking a closer look at the firms' heterogeneity, we selected some firms with a stereotypical development. In Figure 7.4 we see what intuitively seems to be obvious: the earlier a firm enters, the better it is able to increase its market share. All the same, there is no guarantee that first-movers will survive. The first-mover in Figure 7.4 dominates the market until it is outperformed by the two successive firms and eventually has to exit the market because of insolvency caused by an unbalanced set of endowments within the firm. Best practice is shown by firm 5, which is a relatively late entrant but dominates the market until the end of the simulation run. The performance, that is, the growth rate, the size and the time span of survival, depend on a firm's set of endowment.

This should suffice for a rough description of the endogenous development of the sector as we do not discuss market structure and firm size.

Figure 7.5 presents a cognitive argument which we consider the guiding element of entrepreneurial behavior. Actors have to evaluate their chances of founding a potentially successful firm. Due to their bounded rationality, they have to decide on the grounds of their accumulated knowledge and

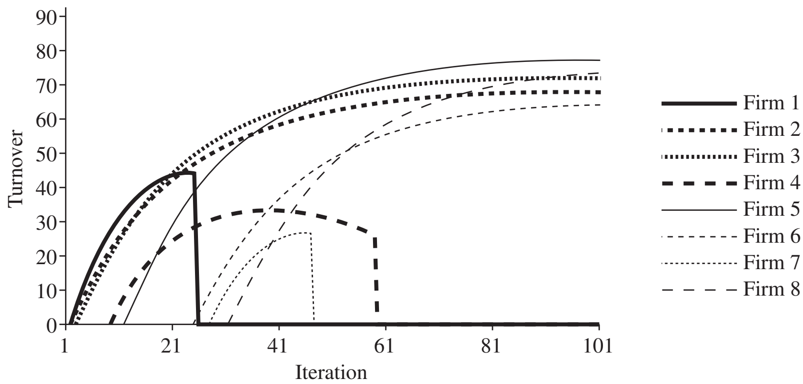


Figure 7.4 Turnover of firms within the sector

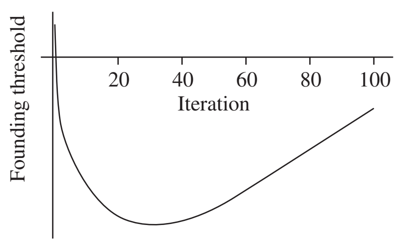


Figure 7.5 Founding threshold

experiences whether to found a firm or not. They make a subjective decision which is influenced by their perception of market opportunities, represented by the individuals' interpretation of the economic indicators. The higher a sector's growth rates, the better market opportunities are evaluated. Hence the actors' inhibition about entrepreneurial behavior decreases and more and more firms are founded. Thus the foundation threshold decreases until compensating effects set in: with an increasing number of firms in the market, the competitive thread is increased. Furthermore, growth rates shrink and some firms already have to exit the market. As economic indicators get worse, the foundation threshold starts to rise. Correspondingly, we observe a swarm of entrepreneurs (Figure 7.6) along the plummeting foundation threshold and a decreasing number of firm

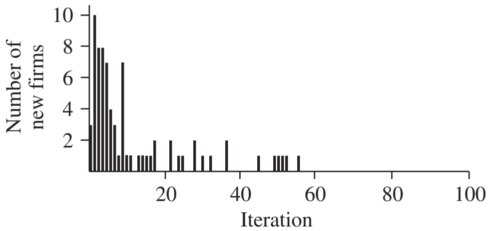


Figure 7.6 *Number of new firms per period*

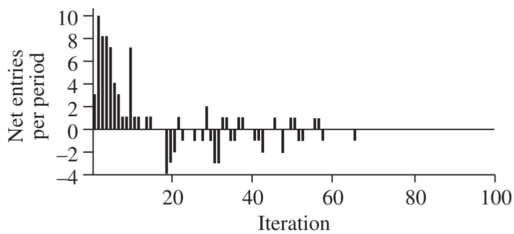


Figure 7.7 *Net entries per period*

entries when exits occur first. The foundation threshold starts to rise again. Fewer actors evaluate market opportunities positively and found a firm. Exits exceed entries (Figure 7.7).

Since we have only introduced the sector's growth rate in turnover and the number of exits as two of many other measurable economic indicators influencing the actors' perception of the real economic situation, the founding threshold keeps on rising after plummeting once. Therefore the formation of new firms fades out gradually. So do exits, as we have not yet implemented the long-run selection process of 'competition in the market'. This shows that there are still some chores to be done in our future research work.

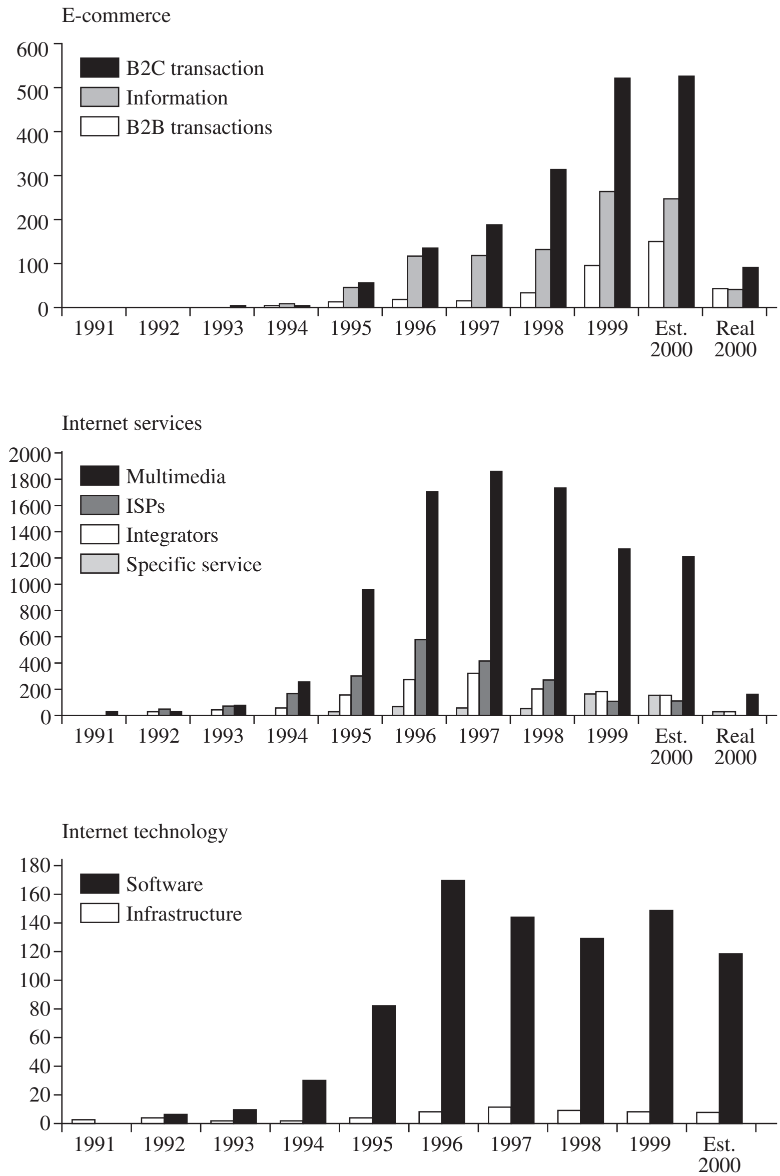
6. SOME EMPIRICAL FINDINGS

The formulation of the model and the first simulation results delivered the intuition and the functioning of the model. In a next step an empirical validation has to be undertaken to round off the analysis. It has already been stated that the model is to serve as a platform approach to be calibrated and possibly reformulated to achieve robustness of the model's implied hypotheses. The construction process of the model itself was inspired by various empirical works such as Klandt (1984), Szyperski and Nathusius (1977) and Brüderl et al. (1996). The endowment set of actors summarizes all possible characteristics of the individuals which might have an influence on entrepreneurial behavior. The role of social networks in a pre-entrepreneurial phase has been discussed by Birley (1985). Klandt and Krafft (2001) investigated the foundation of Internet/e-commerce firms in Germany, surveying 8989 newly founded firms via an online questionnaire.[47] They state that on average 1.9 (in firms not financed by venture capital, Business Angels or strategic investors) to 3.1 (in firms financed by venture capital, Business Angels or strategic investors) individuals take part in a foundation. Furthermore, the analysis of 1890 start-up firms delivered the results depicted in Figure 7.8.

In each sector a wave of firm foundations shows up. The first wave was in the technology sector, followed by Internet services and then e-commerce. Figure 7.9 shows insolvencies of Internet/e-commerce firms per month, where we can see a surging number of exits following the swarms of foundations. Venture capital has a significant influence on the firms' growth. In 1999, start-up firms financed by venture capital generated sales of 2.6 million DM on average, whereas start-up firms without venture capital came up to only 1.4 million DM.[48] Aside from this, the propensity to insolvency is higher among venture-capital-financed firms (Figure 7.10) Although we have not yet included a differentiation between venture-capital-financed and non-venture-capital-financed firms,[49] the intuition that a high amount of venture capital increases the propensity to insolvency can be corroborated. Concerning the actors' attitude towards innovative technologies such as the Internet, we cannot yet offer an empirical validation of the so-called founding threshold, which represents the dynamic change of the actors' evaluation of market opportunities, contingent to the underlying feedback processes we assumed in our model. This will be left for future research.

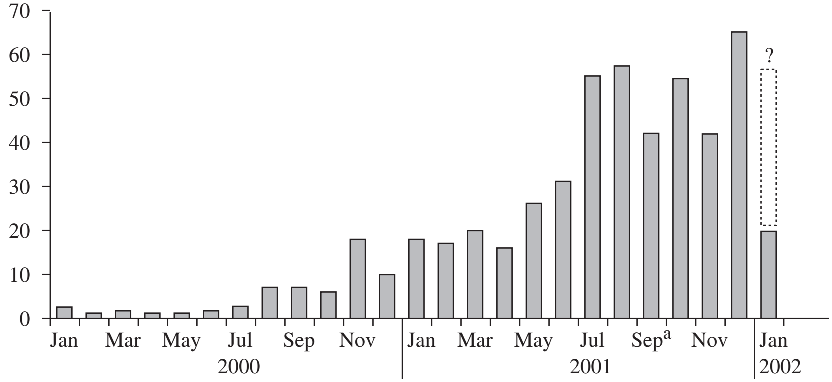
7. CONCLUDING REMARKS AND FURTHER RESEARCH

We developed a model of entrepreneurial behavior which we claim to be an evolutionary one. Once again, we emphasize that we explicitly consider entrepreneurial behavior – the birth process of firms and industries. A further discussion of the industry life cycle is not intended. The core elements of the model are the heterogeneity of actors, their bounded rational



Source: e-startup.org database, Newsfeeds, RWS-Verlag (www.rws-verlag.de/inda/inso.htm), Insolne GmbH (www.insolnet.de)

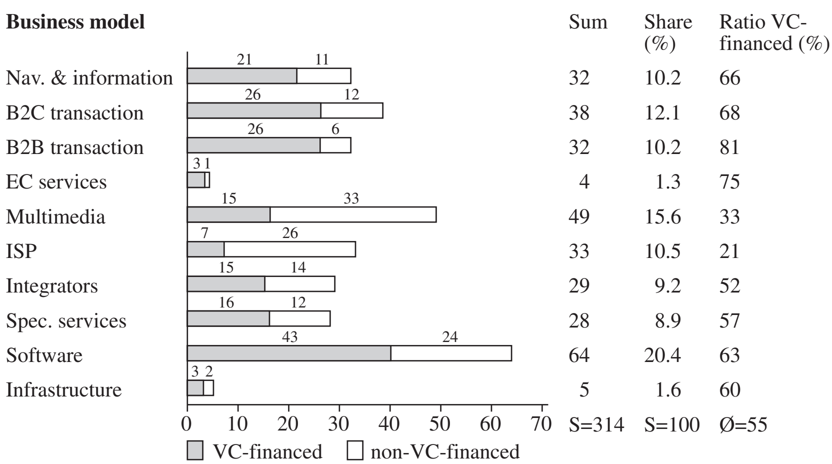
Figure 7.8 Swarms of firm foundations



Note:
[a] The numbers for September are lower than depicted in the diagram because of a lag between acceptance and publication of insolvencies.

Source: e-startup.org database, Newsfeeds, RWS-Verlag (www.rws-verlag.de/inda/inso.htm), Insolne GmbH (www.insolnet.de).

Figure 7.9 Insolvencies of Internet/e-commerce firms per month



Source: e-startup.org database, Newsfeeds, RWS-Verlag (www.rws-verlag.de/inda/inso.htm), Insolne GmbH (www.insolnet.de).

Figure 7.10 Insolvencies by business model: VC-financed vs non-VC-financed behavior to make myopic decisions in favor of founding a firm (which might eventually lead to suboptimal outcomes), the feedback effects from the micro- to the macro-level and vice versa, the (irreversible) historicity of events and the variation and selection mechanisms that put the economic process into a dynamic context.

Neither using an equilibrium concept nor assuming optimal behavior, we manage to avoid a 'survivor bias', at least from a theoretical point of view. Some actors decide to run a firm even though they have to exit in the short run because of a lack of the necessary and adequate comprehensive endowment.

Economic change is brought about, first, by the actual economic development driven by the market process and, second, by the changing attitude of actors driven by their perception of the economic situation.

At the beginning of the up-coming new sector, actors have to deal with true uncertainty prevailing in the decision-making process. They have to rely on their subjective and possibly 'false' intuition concerning their entrepreneurial actions, which leads to market turbulence in the early phase of the sector's life cycle. As time goes by, actors are better able to understand new technologies, to assess market opportunities and their chances of a successful innovative, entrepreneurial behavior. Consequently, uncertainty decreases. More precise predictions and more careful decisions will be made so that stabilizing forces set in.

Our future research work is motivated by empirical applications. Therefore, some specifications will be necessary. Starting at the actors' level, we have to investigate the actors' individual set of endowments in order to identify the actual essential components that spur entrepreneurial behavior, including the creative process of generating a business idea. Aside from this, a possible classification of actors and the formation process of their social networks that have an impact on entrepreneurial behavior have to be considered. In this context, we will have to introduce an interaction-based component into our model to illustrate the qualities of the actors' search process.

The most challenging part of our future research work will be to analyze the cognitive part of the story, which is the role of the founding threshold. It is to investigate the way economic actors perceive the economic situation and a universal mental construct comes into existence, leading to a bandwagon effect in entrepreneurial actions and showing swarms of innovations.

NOTES

1. Barreto (1989).
2. Cantillon (1931).
3. See Hébert and Link (1982) for an overview.

4. Knight (1921).
5. Quesnay (1888).
6. Hébert and Link (1982, p. 31).
7. Baudeau (1919).
8. Turgot (1977).
9. Say (1840 and 1845).
10. Distinguishing Say's concept from Cantillon's.
11. Hébert and Link (1982).
12. See ibid.
13. Barreto (1989).
14. Walras was certainly not the only influence on Schumpeter's thinking. Many others delivered preparatory work, such as Marx, Weber, Menger, Wieser, Say, Hayek and Böhm-Bawerk, to name a few. But as the equilibrium concept is the bone of contention we quote Walras in this context. See Hébert and Link (1982) for a brief overview.
15. Although Schumpeter was fascinated by Walras's concept of equilibrium, the bifurcation point of their intellectual paths originated in the different treatment of the subject. Walras thought it was permissible to abstrahize beyond the adjustment processes in an economic system starting right at the end, which is equilibrium (see Bürgermeister 1994). Schumpeter concentrated more on the process that destroys equilibrium and, if at all, might lead to equilibrium.
16. Barreto (1989).
17. Ludwig von Mises and Friedrich von Hayek as the alleged leaders of the Austrian school engaged in the analysis of disequilibrium conditions focusing on market processes. To get a good idea of Hayek's attitude towards mainstream economics, see Hayek (1937). Concerning Ludwig von Mises, some necessary amendments will be given when introducing Kirzner's entrepreneur later on in this chapter.
18. This is the point at which to stress the role of information and knowledge, as Hayek, Mises and Kirzner do.
19. Kirzner (1973, p. 10).
20. Robbins puts forward the economic agents' task to efficiently economize on scarce resources. But efficiency is no more possible in an Austrian-school-like market process (Robbins 1962).
21. Von Mises (1959).
22. Barreto (1989, p. 17).
23. Ibid., p. 21.
24. Kirzner (1999).
25. Hodgson (2000).
26. For a succinct setting of an evolutionary theory, see, for example Nelson (1995).
27. As the reference work on uncertainty see Knight (1921) and his distinction between risk and uncertainty. In the entrepreneurial context we have to deal with 'true' uncertainty. The agent neither knows the outcome nor is able to calculate corresponding probabilities.
28. For this discussion see, for example, Simon and Egidi (1992).
29. By personality we mean the conglomerate of accumulated knowledge, information and experience.
30. Each component is the result of a cumulative evolutionary process which will not be discussed in this chapter. With respect to an empirical application, each component requires sector-specific observations.
31. Schlutz (1971).
32. Becker (1993).
33. We are conscious of our tightrope walk in using a strictly neoclassical concept within our model, which we explicitly claim to be evolutionary. We assume a link between the agents' set of capabilities and their economic performance. For the time being, we use it as a metaphor to stress the importance of knowledge in our model, leaving the necessary 'evolutionary' clarification of this concept for further research.
34. Hayek (1937).

35. Loasby (1999).
36. Schumpeter (1939).
37. Blanchflower and Oswald (1995).
38. Besides the suggested endowment factors, any other desired factor can be included into the endowment set.
39. See Blanchflower and Oswald (1995).
40. As we proceed we will not confine the entrepreneurial behavior to a single agent but to a number of agents.
41. This accords with Schumpeter (1939).
42. We could include learning in the model and thus reflect the human capital component. Coase (1988), Penrose (1959), Demsetz (1973), Wernerfelt (1984) definitely give enough inspiration to extent our model.
43. Granovetter (1973).
44. See Aldrich and Wiedenmayer (1993) as an example.
45. Hall (1982).
46. In the event agents do not have a sufficient set of endowments and, hence, need additional resources, complementary assets and competencies, networking plays an important and manifold role. Not only does the social network provide the opportunity to have access to additional and complementary endowment factors, but networks also have a crucial influence on the actual entrepreneurial decision to start a venture itself. Suppose a single agent thinks himself unable to start a business all by himself. Then he has to persuade others to support him. Otherwise, the lack of legitimacy may prevent entrepreneurial actions. Whereas a high degree of innovativeness, the so-called liability of newness, might be ended by an agent's obstructive social network, a synergetic outcome of either strong or weak ties within a network can be an enhanced and by the group subjectively high-valued business idea. In other words, a social network functions as a catalyst sparking a venture. See Efring and Baden-Fuller (2000).
47. For more details see also www.e-startup.org.
48. Compare Klandt and Krafft (2000).
49. All actors in our model have a certain amount of 'venture capital', that is, freely disposable money capital. So there is no such differentiation between venture-capital-financed vs non-venture-capital-financed firms. None the less, once we incorporate different populations in our actors' base, including a population of venture capitalists, and furthermore work on a proper representation of a search process that brings the appropriate actors together, the model will deliver corresponding results.

REFERENCES

Aldrich, H.E. and Wiedenmayer, G. (1993), 'From traits to rates: an ecological perspective on organizational foundings', in R.H. Brockhaus and J. Katz (eds), *Advances in Entrepreneurship, Firm Emergence, and Growth*, Greenwich, CT: JAI Press, pp. 145–95.

Barreto, H. (1989), *The Entrepreneur in Microeconomic Theor*y, London and New York: Routledge.

Baudeau, N. (1919), 'Première introduction à la philosophie économique', Paris: Geuthner.

Becker, G.S. (1993), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Chicago: University of Chicago Press.

Birley, S. (1985), 'The role of networks in the entrepreneurial process', *Journal of Business Venturing*, **1**, 107–17.

Blanchflower, D.G. and Oswald, A. (1995), 'What makes an entrepreneur?', unpublished.

Bürgermeister, B. (1994), 'The misperception of Walras', *American Economic Review*, **84** (1), 342–52.
Brüderl, J., Preisendörfer, P. and Ziegler, R. (1996), *Der Erfolg neugegründeter Betriebe – eine empirische Studie zu den Chancen und Risiken von Unternehmensgründungen*, Berlin: Duncker & Humblot.
Cantillon, R. (1931), *Essai sur la nature du commerce en général*, H. Higgs.
Coase, R.H. (1988), *The Firm, the Market, and the Law*, Chicago: University of Chicago Press.
Demsetz, H. (1973), 'The theory of the firm revisited', *Journal of Law und Economic Organization*, **4**, 151–62.
Efring, T. and Baden-Fuller, C. (2000), 'The locus of entrepreneurship: Firms, networks and markets', unpublished.
Foss, N.J. (1993), 'Theories of the firm: Contractual and competence perspectives', *Journal of Evolutionary Economics*, **3**, 127–44.
Granovetter, M.S. (1973), 'The strength of weak ties', *American Journal of Sociology*, **78**, 1360–80.
Hall, R. (1982), *Organizations: Structure and Process*, 3rd edn, Englewood Cliffs, NJ: Prentice-Hall.
Hayek, F.A. (1937), 'Economics and knowledge', in M. Casson (ed.), *Entrepreneurship*, Aldershot: Edward Elgar, reprint 1990.
Hébert, R.F. and Link, A.N. (1982), *The Entrepreneur: Mainstream Views and Radical Critiques*, 2nd edn, New York: Praeger.
Hodgson, G.M. (2000), 'What is the essence of institutional economics?', *Journal of Economic Issues*, **34** (2), 317–30.
Jolink, A. (1996), *The Evolutionist Economics of Léon Walras*, London and New York: Routledge.
Kirzner, I.M. (1973), *Competition and Entrepreneurship*, Chicago: University of Chicago Press.
Kirzner, I.M. (1999), 'Creativity and/or alertness: A reconsideration of the Schumpeterian entrepreneur', *Review of Austrian Economics*, **11** (1/2), Kluwer Academic Publishers, pp. 5–17.
Klandt, H. (1984), *Aktivität und Erfolg des Unternehmensgründers: Eine empirische Analyse unter Einbeziehung des mikrosozialen Umfeldes*, Bergisch Gladbach.
Klandt, H. and Krafft, L. (2000), *Internet/E-Commerce Gründungen in Deutschland, Segment-Analyse: Verkaufsorientierte B2C-Anbieter (Business-to-Consumer)*, Diskussionspapier. Stiftungslehrstuhl für Gründungsmanagement und Entrepreneurship, European Business School.
Klandt, H. and Krafft, L. (2001), *Aktuelle Beschäftignung und Mitarbeiterbedarf bei Internet-/E-Commerce-Gründungen in Deutschland*, Diskussionspapier, Stiftungslehrstuhl für Gründungsmanagement und Entrepreneurship, European Business School.
Knight, F.H. (1921), *Risk, Uncertainty and Profit*, New York: Houghton Mifflin.
Loasby, B. J. (1999), *Knowledge, Institutions and Evolution in Economics*, New York: Routledge.
Nelson, R.R. (1995), 'Recent evolutionary theorizing about economic change', *Journal of Economic Literature*, **33**, 48–90.
Penrose, E.T. (1959), *The Theory of the Growth of the Firm*, Oxford: Oxford University Press.
Quesnay, F. (1888), *Oeuvres économiques et philosophiques*, Frankfurt: M.J. Baer.

Robbins, L. (1962), *An Essay on the Nature and Significance of Economic Science*, 2nd edn, London: Macmillan.
Say, J.B. (1840), *Cours complet d'économie politique pratique*, Paris: Buillamin.
Say, J.B. (1845), *A Treatise on Political Economy*, 4th edn, Philadelphia: Grigg & Elliot.
Schlutz, T.W. (1971), *Investment in Human Capital*, New York: Free Press.
Schlutz, T.W. (1975), 'The value of the ability to deal with disequilibria', *Journal of Economic Literature*, **13**, 827–46.
Schumpeter, J.A. (1911), *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, Cambridge: Harvard University Press, translated by Redvers Opie in 1934.
Schumpeter, J.A. (1939), *Business Cycles*, New York: McGraw-Hill.
Simon, H.A. and Egidi, M. (1992), *Economics, Bounded Rationality and the Cognitive Revolution*, Aldershot; UK and Brookfield, USA: Edward Elgar.
Szyperski, N. and Nathusius, K. (1977), 'Gründungsmotive und Gründungsvorbehalte – Ergebnisse einer empirischen Studie über potentielle und tatsächliche Unternehmensgründer', *Die Betriebswirtschaft*, Vol. 37, Jena, pp. 299–309.
Turgot, A.R.J. (1977), *The economics of A. R. J. Turgot*, The Hague: Martinus Nijhoff.
von Mises, L. (1959), *Human Action*, London: William Hodge.
Wernerfelt, B. (1984), 'A resource-based view of the firm', *Journal of Strategic Management*, **5**, 171–80.

8. Shakeout in industrial dynamics: new developments, new puzzles

Jackie Krafft

1. INTRODUCTION

The growing body of analysis in the field of industrial dynamics since the 1980s may lead people to think that industrial dynamics is a new domain of research. This is of course a misperception since some early contributions provided first steps towards the elaboration of an industrial dynamics approach. Schumpeter (1912, 1942) did significant work emphasizing the role of the entrepreneur in the development of innovation, as well as the evolution of industry in a context of radical change. Marshall (1890, 1920) also proposed many lines of enquiry, such as the fact that the economy is composed of different sectors, the growth and decline of which are unequal and intrinsically dependent on the organization of knowledge.

This misperception is certainly due to the fact that the post-Second World War decades were characterized by the development of industrial organization that focused on optimality properties and comparative efficiency studies of different equilibrium situations, and ignored the conditions under which an industry could emerge and evolve over time. From the late 1950s to the 1960s, the structure–conduct–performance paradigm, at the core of the Harvard tradition, focused essentially on the determinants of features of market structure and performance, such as concentration, firms' size and profitability. In the 1970s and 1980s, more attention was given to how the behaviour of firms could have an impact on market structures and performances, with the emergence of new approaches such as the Chicago School, the Theory of Contestable Markets and the New Industrial Organization. These approaches were major advances in the development of conventional industrial organization. But in retrospect they made meagre contributions to the specific problem of industrial dynamics. They were not based on a dynamic framework but on a static one, and their unit of analysis is not the industry but the rational behaviour of firms.

Over the 1980s, however, some authors tried to build on the neglected work of Schumpeter and Marshall, and focused on major changes that have taken place in industry structure, industrial leadership, economic growth and innovation. The research programme initiated by Nelson and Winter (1982), which focused on evolutionary theory and economic change, opened the door to new interpretations. In one of these new interpretations, Gort and Klepper (1982) tried to understand the long-term evolution of innovative industries, and assessed that this long-term evolution is essentially characterized by a life cycle in which industries, like bio-organisms, arise in their birth time, grow and mature in their development time, and decline in their death time. The industry life cycle clearly added value to the explanation of a large number of regularities occurring in innovative industries. But the shakeout, which corresponds to a massive exit of producers, progressively became a central regularity to be explored in industrial dynamics. Most of the recent debates have attempted to clarify when and why a shakeout occurs.

We can first think that technology essentially drives the life cycle of an industry, and is responsible for the shakeout. This calls to mind Schumpeter's vision of creative destruction in industrial dynamics. An entrepreneur sets up a firm to introduce his invention. This firm grows and holds a monopoly position for some time. But in time this firm is imitated by new entrants which compete severely, and eventually outperform the initial firm. This situation can last until another entrepreneur develops a new project involving the exit of older and larger firms and the entry of new ones.

But we can also think about the shakeout in a different manner. We can consider that knowledge and competencies drive the life cycle of the industry. In that case, closer to Marshall's vision, the growth of knowledge is linked to the ability of firms to ensure a coherence between internal economies (organization and direction of the resources of the firm) and external economies (general development of the economy, including the role of firms in the neighbourhood). In this perspective, the shakeout affects firms differently, since some firms might have the opportunity to accumulate specific knowledge and competencies, and survive. In some cases non-shakeout patterns may thus emerge.

The purpose of this chapter is to stress these different visions of shakeout in industrial dynamics, to characterize the new developments that support these alternative visions, and to clarify the puzzles that these visions may present for industry life cycle analyses. The next section provides a synthesis of the basic framework of the industry life cycle proposed by Gort and Klepper in the 1980s. The remainder of the chapter investigates more recent developments on shakeout, which is considered a central question in the 1990s and 2000s. The third section focuses on a conception of shakeout which is closely linked to technological conditions. The obsolescence of an old technology and the dominance of a new one involves a new industry replacing the old one. Firms in the old industry have to face a shakeout, and these firms are generally not the main actors of the new industry. The fourth section focuses on an alternative conception of shakeout which is closely connected to knowledge accumulation and diffusion. A new industry never starts from scratch, as industries generally arise through a transformation of existing industries. When a shakeout occurs, firms that organized the conditions of knowledge accumulation and diffusion may survive, and eventually become the leaders of the newly born industry. The fifth section offers some concluding remarks.

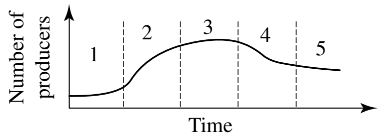
2. INDUSTRY LIFE CYCLES

The main ambition of Gort and Klepper (1982) is well known. The study attempts to measure and analyse the diffusion of product innovations and views the historical sequence or time path of events as a critical determinant of the ultimate structure of new product markets. In this perspective, the study addresses a series of new questions: how does innovation proceed over time and affect the structure of the industry? What explains the fact that information either favours or blocks entry throughout the process of innovation? What are the main regularities that drive the emergence, maturity and decline of innovative industries? Let us examine these different questions in turn.

2.1 Diffusion of Innovation and Evolution of Industry

The empirical part of the chapter focuses on 46 product histories and describes how these 46 new product innovations are diffused in the market. These product histories serve as the basis for a theory of the development of industries for new products.

There are five different stages in product histories, which structure the development of industries and characterize the life cycle of an industry (Figure 8.1). The first stage begins with the commercial introduction of the new product by its inventor or first producer. The size of the market is very limited, the commercial success of the product is highly uncertain, and the product is a kind of prototype to be further improved. This first period ends when new entrants start to penetrate the newly born industry. The length of this stage depends on the ease of copying the initial innovator, the size of the market for the new product after its introduction, and the number of potential entrants into the market. The second stage is characterized by an increase in the number of producers of the new product. Growth of output



The five stages of evolution are determined by the following process:

$F_t = P_t\,(N - n_{t-1})$

with F_t, the expected number of entrants in t;
P_t, the probability of entry in t of each potential entrant;
N, the population of potential entrants;
n_{t-1}, the number of firms that have already entered the market by $t-1$.

Figure 8.1 The five stages of new product industries

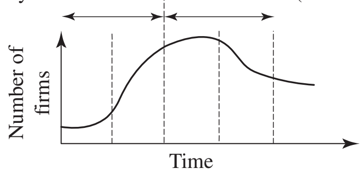
is high and the design of the product is narrowly specified. The third stage corresponds to a net entry equal to zero (the number of entrants is roughly balanced by the number of exiting firms). Process innovation progressively replaces product innovation, since the process of production is more complex and has to be oriented towards a lower-cost development of the innovation. This stage ends with a decline in gross entry. In the fourth stage, there is a negative net entry, and the number of exiting firms is far higher than the number of new entrants. Finally, a large number of incumbent firms disappear, exiting the industry in the fifth stage. This phenomenon – called the shakeout – corresponds to the maturity of the industry.

2.2 Types of Information and Entry/Exit Process

Entry is defined by a probability P_t and evolves from stage to stage according to different innovation behaviours. Innovation behaviours depend on the access to information, as well as on profit opportunities (Figure 8.2). In stage 2 innovations essentially come from external sources of information. This first type of information is accessible to any potential new entrant and favours the process of entry, as large opportunities of profit are available. However, in stages 3 to 5, information is now more related to the experience accumulated by incumbent firms on both the nature of the new product and the process of production. This second type of information acts as a barrier to entry. Profit opportunities are more reduced for potential entrants and the process of entry stagnates progressively.

In these different stages, innovation is not considered an isolated phenomenon generating a new technological trajectory. Rather, innovation appears all along the life cycle and the nature of innovation changes as the life cycle progresses. Product innovation at the beginning of a new cycle has a major impact on production costs and product quality. Process innova-

Stages 1 to 2:	**Stages 3 to 5:**
Information (external)	Experience (internal)
Product innovation (radical)	Process innovation (incremental)
Massive entry	Massive exit (shakeout)



Entry probability P_t is characterized by the following function:

$P_t = f\,(I_{2t}, L_t, \Pi_t)$

with I_{2t}, number of innovations at time t emanating from sources external to the industry;
L_t, the accumulated stock of experience of incumbent producers;
Π_t, profit of incumbent producers at time t;
and $\delta f/\delta I_{2t} > 0$, $\delta f/\delta L_t < 0$ and $\delta f/\delta \Pi_t > 0$.

Figure 8.2 Information, innovation and entry/exit process

tion at the end of the life cycle involves more incremental adjustments on production processes and management/marketing techniques.

2.2.1 Regularities in emerging industries

The industry life cycle is governed by the existence of six regularities or principles of evolution (Figure 8.3):

- production increases in the initial stages and declines in the final stages;
- entry is dominant in the early phases of the life cycle and is progressively dominated by exit. A massive process of exit (a shakeout) occurs in the final stages of the life cycle;
- market shares are highly volatile in the beginning, and tend to stabilize over time;
- product innovation tends to be replaced by process innovation;
- first movers generally have a leadership position which guarantees their long-term viability;
- product variety disappears over time, as a dominant design emerges.

3. SHAKEOUT AND THE EVOLUTION OF TECHNOLOGY

The initial project of describing industry life cycles was to explore a large number of regularities in the development of industries. In the 1990s,

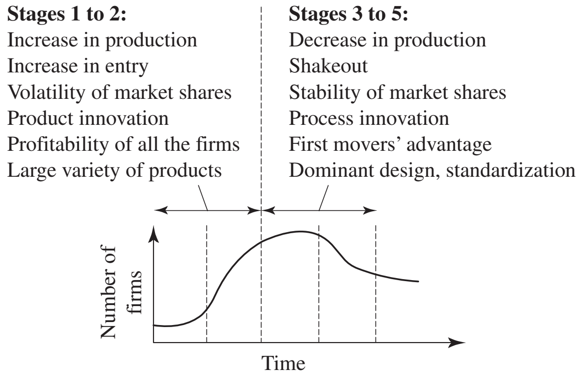


Figure 8.3 Regularities in the development of industries

however, the project focused more and more on the shakeout phenomenon, and attempted to clarify what occurs in pre-shakeout versus post-shakeout periods. This attention is of course related to the crucial role of shakeout in the industry life cycle: a cycle cannot be observed without a shakeout in the mature stages of the industry. But shakeout is also a key to understanding why a given industry is declining, and why major actors of this industry tend to be superseded by new actors creating a new industry. Behind this, there is the idea that a given technology can create profit opportunities for some time, but that new technologies will recurrently be created and replace older ones. This Schumpeterian vision of the dynamics of an economic system has been explored in recent contributions on the shakeout in industry life cycle, with an emphasis on different determinants. A series of empirical results has also been completed to validate the predominance of a life cycle in innovative industries.

3.1 Shakeout and Dominant Design

Abernathy and Utterback (1978), Clark (1985), and Utterback and Suarez (1993) develop an analysis of shakeout which is derived from the traditional Schumpeterian hypothesis on the R&D advantage of large firms. Large firms are generally engaged in important R&D programmes which generate new products. When a large firm selects one of these new products and decides to launch it on the market, this large firm must face a high level of uncertainty affecting the conditions of both demand and supply. On the demand side, uncertainty comes from the fact that the firm does not know the details of customers' preferences, that is, preferences related to the various possible characteristics of the product innovation. On the supply side, the conditions of production are also highly uncertain and may evolve over time. Different producers can thus try various product innovations with distinct characteristics, and implement different processes of production. These alternative producers thus engage in a process of competition.

Over time, however, uncertainty decreases and selection operates. On the demand side, uncertainty decreases once customers of the new product have tested the alternative characteristics, and acquired experience of what they expect from the new product, and which characteristics are more adapted to their personal taste and usage. Eventually customers select a series of product characteristics and demand becomes more predictable. On the supply side, rival producers learn over time and accumulate information on what customers prefer. In time they also select a series of production techniques which are adapted to low-cost production.

Since uncertainty decreases, the shakeout appears as an endogenous phenomenon. Product innovation diminishes because most of the actors (producers and customers) are naturally oriented towards the production and consumption of a standardized good. The progressive emergence of a dominant design involves higher barriers to entry which correspond to investments by incumbents in process innovation. Entry is thus limited, and less efficient incumbent firms exit the industry.

3.1.1 Shakeout and technological shock

Jovanovic and McDonald (1994) propose a very different vision of shakeout. For these authors shakeout is generated by an external technological shock, exogenous to the industry. The first technological shock sets in with the development of the new product being launched on the market. Entry is stimulated by the emergence of new profit opportunities related to this new technology/new product, but subsequently there is a progressive reduction in profit margins and the industrial structure stabilizes on a limited number of firms in the industry. At this stage, which corresponds to the maturity of the industry, a new technological trajectory emerges and again stimulates the process of entry, in the meantime, involving an adjustment of incumbent firms. The process of adjustment is driven by a stochastic process and only a few firms survive this external shock. The shakeout thus eliminates firms that failed to adapt themselves to the new technology.

3.1.2 Shakeout and timing of entry

Finally, Klepper (1996) relates the shakeout to the timing of entry. The reference is, here again, the Schumpeterian hypothesis on the relation between firms' size and R&D capacity. But the novelty is that this hypothesis is discussed on the basis of a finer distinction between firms which can eventually be incumbent, new entrant, or latecomer.

Process innovation reduces the average costs of large firms, which are the major actors in this type of innovation. However, some key elements may erode the advantage of larger firms. For instance, large firms have to cover specific costs, such as expansion costs, which limit their growth. The activity of R&D can also exhibit decreasing returns to scale over time. Because of these elements, early entrants can develop process innovations, sometimes much better than incumbents or latecomers. Early entrants can thus enjoy a leadership position in process innovation as, on the one hand, incumbents have to deal with other problems related to their large size and, on the other hand, latecomers have to concentrate on product innovation, which allows them to grow to a minimum size in order to survive. The timing of entry is thus a major determinant in the formation of a competitive advantage over incumbents, as well as in long-term survival over latecomers. This mechanism provides an alternative explanation of the shakeout.

3.1.3 Empirical studies on shakeout

In their pioneering work, Gort and Klepper (1982) considered 46 different industries and showed that they evolved according to a life cycle. More recently, Agarwal (1998) completed the time series until 1991, and confirmed that the industries that were in their maturity stage in the early 1980s faced a shakeout within the decade. The initial programme, which was based on research of large empirical validation, appears to have been maintained today (see also Klepper and Graddy, 1990).

However, most of the contributions in the late 1990s tend to determine a limited set of industries whose evolution largely conforms to the industry life cycle. Some of the 46 initial industries therefore deserve special attention: cars, typewriters, car tyres, commercial aircraft for trunk carriers, televisions, television picture tubes, and penicillin (Klepper 1997, Klepper and Simons 1997, 1999, Klepper, 2002). And shakeout is of course one of the main regularities to be investigated empirically within these industries.

An important part of the investigation concerns the definition of shakeout. In most cases, shakeout was considered a massive exit occurring in the maturity stage, which involved negative net entry rates. Klepper and Miller (1995) clarify this definition. They stress that a product is deemed not to have experienced a shakeout if the number of firms never declines below 70 per cent of the peak number, or if it does but subsequently recovers to over 90 per cent of the peak.

Another part of the research concerns the determination of the most dominant explanation of shakeout in these industries (Klepper and Simons 1999). The different theses on shakeout – shakeout and dominant design, shakeout and technological shock, shakeout and timing of entry – are thus competing with each other, and the following results emerge from this empirical confrontation. The more a firm ages in the industry, the lesser its probability of exit. As a matter of fact, firms which penetrated the industry before the shakeout, and thus compose the cohorts of pre-shakeout, have the lowest exit rates when the shakeout occurs. Green et al. (1995) show that most of these firms capture large market shares and produce a large spectrum of products. On the contrary, cohorts of post-shakeout which regroup younger firms and latecomers are characterized by the highest exit rates. Some of them survive (Agarwal and Gort 1996), but occupy small market niches (Klepper 1997). It appears, moreover, that the first movers' advantage which benefits the cohorts of pre-shakeout is directly connected to their capacity of adjustment in terms of innovative behaviour (from product innovation to process innovation). Shakeout is thus essentially driven by timing of entry, and to a lesser degree by alternative explanations which include the emergence of a dominant design or a new technological trajectory.

4. SHAKEOUT AND EVOLUTION OF KNOWLEDGE

The idea of a shakeout essentially driven by the evolution of technology over the course of the industry life cycle is progressively challenged by a new vision. Henderson (1995) shows that the traditional technology of optical lithography used in the semi-conductor industry should have been replaced by new and superior technologies such as X-rays or electrons. This traditional technology nevertheless persisted and had an unexpected long old age since the final users continued to privilege this technology, and a specific organization of the industry was implemented, with firms developing complementary competencies. Moreover, Mueller (1990, 1997) shows that the long-term viability of first movers is related in a large number of industries to specific features of demand (such as set-up and switching costs, network externalities of final users, inertia effects due to the customer's uncertainty about quality, inertia effects due to the customer's experience of existing products and services), as well as supply (such as set-up and network externalities of producers, economies of scale, cost-reducing learning by using). Finally, Van Dijk (1998) shows that increasing returns in R&D is not the major element in the first movers' competitiveness, but that network effects have a quite decisive effect. The development, accumulation, diffusion and usage of competencies are thus key elements which drive the industry life cycle and, as an outcome, involve a noticeably different vision of shakeout which is closer to the Marshallian tradition.

4.1 Shakeout and Complementary Competencies

Industry life cycle analyses generally focus on industries in which competition and innovation proceed from the interaction between firms (incumbents and entrants) within a given market, delimited by the purchases and sales of a homogeneous product. On some occasions, however, vertical relationships between firms in the industry and their direct suppliers or customers have a strong impact on the evolution of industries. Innovation processes require the accumulation of complementary competencies, as well as effective coordination between firms which generate these competencies. Since the industry is characterized by strong coordination between suppliers and producers, or producers and retailers, processes of entry and exit become industry-specific. Alternative life cycle patterns thus appear, eventually with non-shakeout phenomena.

In some industries the emergence of specialized suppliers tends to re-dynamize the entry process in the phase of maturity. They develop new production processes, new specialized equipment, new technology at the upstream level and sell it to any downstream potential entrant who can pay the price. They significantly reduce barriers to entry and favour competition. This mechanism of accumulation and transfer of new competencies is observed and documented in various industries.

In chemicals and petrochemicals, production increased rapidly after the Second World War in the USA. After 20 years of successful commercialization of various products (styrene, synthetic fibres and plastics), a large number of specialized engineering firms emerged to service the producers. These specialized suppliers developed new methods of production which could be acquired by both incumbents and new entrants in petrochemicals via patents and licences. Thus, although the US chemical industry has enjoyed an early advantage in petrochemicals, this advantage has been progressively eroded as technologies diffused, enabling Europe and Japan to narrow the gap and other countries to enter the industry. Increased competition, compounded by the oil shock of the 1970s and slowing possibilities for significant product innovations, has culminated in a continuing process of restructuring. The restructuring results in more product-focused firms and more globalized firms. It also generates a reduction in R&D intensity in virtually all subsectors, except the life sciences, which were traditionally a part of the chemical industry but are now a separate industry (Arora et al. 1998).

In telecommunications, liberalization occurred in the mid-1980s in the USA and continued until the late 1990s in Europe. Liberalization is intended to generate a larger spectrum of products and services, with low prices stimulated by intense competition between the incumbents (historical monopolies, generally state-owned) and the new entrants (large diversified groups or small start-ups). A key feature in this period is the fact that incumbents with long experience in R&D within large R&D-dedicated research laboratories decided to delegate this activity to the equipment suppliers located upstream. These equipment suppliers thus have the opportunity to play a major role in the development of new equipments (terminals, networks) which are necessary to exploit new market opportunities (integrated systems for the transfer of voice, data and video, with friendliness, security, reliability and mobility). These equipment suppliers therefore stimulate the entry process of newcomers in the telecommunications industry and for some permit long-term viability (Fransman 1999, Fransman and Krafft 2002).

In other cases, the industry was created by an initial inventor or an academic researcher who decided to set up a firm to exploit the commercial opportunities of his innovation. Often, however, the production and distribution of this innovation required the contribution of other actors, usually larger firms. The coordination of competencies related to innovation on the one hand, and complementary competencies related to production and distribution on the other hand, strongly shaped the profile of evolution of the industry, and stimulated new entries.

In medical instruments, product innovation (X-rays, Nuclear Magnetic Resonance, computed tomographic) is implemented by small start-ups, connected to academic research. These firms are highly specialized and, as the market expands remarkably, they rely more and more on other firms for production and distribution facilities. These other firms are not *de novo* entrants, but rather incumbent producers of prior imaging instruments. Yet with the penetration of the new activity (diagnostic imaging instruments), they significantly extend their competencies. Ultimately, they favour the entry of a new cohort of entrants composed of firms specialized in other complementary competencies such as computed security and testing (Mitchell 1995, Klepper 1997).

4.2 Shakeout and Similar Competencies

The coordination of similar competencies is also an important topic for researchers interested in how innovations occur and their implications for firms and economic change. In some industries there is a somewhat paradoxical phenomenon that both small, specialized firms and large, diversified firms coexist in the long run. Specific firms may come and go, and there are certainly mergers, alliances and bankruptcies, but the two types of firms seem to an extent mutually dependent. This situation may lead to non-shakeout profiles of evolution, with small firms and large firms surviving over the long run.

In biotechnology, networks, clusters and alliances are crucial to the coexistence of similar competencies developed by large and small firms, even though large firms progressively retain control over the networks. It seems important to combine such alliances with in-house R&D and competencies, because otherwise the firm has difficulties in evaluating the potential of new ideas and techniques that are developed outside the firm. In many cases, the intrinsic characteristics of knowledge in terms of codification and appropriability requires extended interaction, and explains why collaboration occurs among firms with similar competencies in order to stimulate innovation. But in the meantime, ownership and control rights are important to understand who has alliances with whom and are absolutely crucial in the evolution of the industry (McKelvey 1996, Saviotti 1998).

4.3 Shakeout and Final Users

When demand is highly diversified, a large spectrum of firms can survive in specific market niches. Their long-term viability is essentially based on their capacity to develop competencies in terms of the products and services customers expect. This greatly favours mechanisms of co-production, co-design and co-innovation, in which customers are highly involved in major stages of development of innovation. As a by-product, non-shakeout patterns can emerge within these industries (Nelson 1998, Windrum and Birchenhall 1998).

In the aircraft industry, for instance, buyers of business jets are generally characterized by the value they place on aspects such as size, speed, distance and manoeuvrability. A large series of aircraft have thus been developed to service these varied needs and firms generally offer only selected aircraft types corresponding to a limited range of buyers. Producers within this submarket are thus highly specialized and clearly engaged in a made-to-measure supply process. This pattern of evolution stands in sharp contrast to the traditional commercial aircraft industry, in which only two firms survived (Boeing and Airbus) after an important and rather traditional shakeout (Klepper 1997). Related aspects on networks and vertical relationships as major non-shakeout determinants emerged in the engine-market segment (Bonaccorsi and Giuri 2001).

In electronics, the emergence of a dominant design, together with a stable oligopoly of producers, is not necessarily achieved. In a sense, the technological convergence which is observed in this industry does not involve product convergence. On the contrary, an increasing process of diversification and market segmentation appears to occur on the basis of a more intricate relationship between the producers and users of the new technology (Gambardella and Torrisi 1998, Ernst 2002).

5. CONCLUDING REMARKS

The notion of shakeout in industrial dynamics has thus stimulated new developments which, in the meantime, involve new puzzles. As a matter of fact, many industries evolve according to the industry life cycle principles. These then face a shakeout when the technology at the origins of the industry is progressively outperformed by a new technology introduced by new entrants. But at the same time many industries do not conform to the industry life cycle framework, either because they are essentially knowledge-driven instead of technology-driven, or because they exhibit non-shakeout patterns of evolution. From what emerges in the previous sections, we would claim that the latter two reasons are not mutually exclusive, but rather connected to some extent. Because of the crucial role of networks, clusters, alliances and cooperations in knowledge-intensive industries, a shakeout does not necessarily occur, or if it does, occurs differently than in technology-driven industries. That is, if the shakeout occurs within these industries, it will affect firms not taking part in a network in which competencies are actively created and coordinated.

Another key puzzle lies in how we define the industry, and which level of aggregation is adequate to characterize a shakeout. The usual procedure is to define the industry on the basis of a specific market in which similar, non-differentiated products are offered by producers and acquired by customers. In knowledge-intensive industries, however, this common procedure only provides a partial outlook on industrial dynamics since suppliers, clients and eventually partners play a crucial role in the evolution profile of the industry. The analysis must thus integrate the role of these actors to determine whether shakeout or non-shakeout patterns are observed in industrial dynamics.

Finally, the current focus on shakeout may eclipse other key problems of industrial dynamics. Industrial dynamics basically analyses the way in which the activities undertaken within the economic system are divided up among firms: some firms embrace many different activities, while for others the range is narrowly circumscribed; some firms are large and others small; some firms are vertically integrated but others are not. Industrial dynamics should therefore describe and clarify how the industry is organized now, how it differs from what it was in earlier periods, as well as what forces were operative in bringing about this reorganization of the industry and how these forces have been changing over time. The study of industrial dynamics demands a permanent and sound connection between facts and theory. The stimulus provided by the patterns, puzzles and anomalies revealed by systematic data gathering and careful collection of detailed information – not only on shakeout but also on other conjectures – is essential to better understanding of the forces that determine the dynamics of industry.

In that respect, the research agenda in industrial dynamics should not be too restricted and should at least continue to proceed along the path opened up by Marshall and Schumpeter (Krafft 2000, 2002). It should further elaborate taxonomies of industrial evolution in order to define some groups of industries that evolve in a similar way (Pavitt 1984, Malerba and Orsenigo 1993, Geroski 1995). The identification of the main relationships between firms, suppliers, customers, competitors and, more generally, governmental, scientific or financial institutions must improve (Malerba and Orsenigo 1996). The link between the evolution of industries, innovation and economic growth, and their implications at the local, national and international level also deserves specific attention (Lazonick 1991, Malerba et al. 1999, Lamoreaux et al. 1998).

REFERENCES

Abernathy, W. and Utterback, J. (1978), 'Patterns of industrial innovation', *Technology Review*, **80**, 41–7.

Agarwal, R. (1998), 'Evolutionary trends of industry variables', *International Journal of Industrial Organization*, **16**, 511–25.

Agarwal, R. and Gort, M. (1996), 'The evolution of market and entry, exit, and survival of firms', *Review of Economics and Statistics*, **78**, 489–98.

Arora, A., Landau, R. and Rosenberg, N. (eds) (1998), *Chemicals and long term economic growth: insights from the chemical industry*, New York: John Wiley and Sons.

Bonaccorsi, A. and Giuri, P. (2001), 'The long term evolution of vertically-related industries', *International Journal of Industrial Organization*, **19**, 1053–83.

Clark, K. (1985), 'The interaction of design hierarchies and market concepts in technological evolution', *Research Policy*, **14**, 235–51.

Ernst, D. (2002), 'Electronics industry', in W. Lazonick (ed.), *IEBM Handbook of economics*, London: Thompson Learning.

Fransman, M. (1999), *Visions of Innovation: the firm and Japan*, Oxford and New York: Oxford University Press.

Fransman, M. and Krafft, J. (2002), 'Telecommunications', in W. Lazonick (ed.), *IEBM Handbook of economics*, London: Thompson Learning.

Gambardella, A. and Torrisi, S. (1998), 'Does technological convergence imply convergence in markets? Evidence from the electronics industry', *Research Policy*, **27**, 445–63.

Geroski, P. (1995), 'What do we know about entry?', *International Journal of Industrial Organization*, **13**, 413–614.

Gort, M. and Klepper, S. (1982), 'Time paths in the diffusion of product innovations', *Economic Journal*, **92**, 630–53.

Green, D., Barclay, D. and Ryans, B. (1995), 'Entry strategy and long-term performance: conceptualisation and empirical examination', *Journal of Marketing*, **59**, 1–16.

Henderson, R. (1995), 'Of life cycles real and imaginary: the unexpected long old age of optical lithography', *Research Policy*, **24**, 631–43.
Jovanovic, B. and McDonald, G. (1994), 'The life cycle of a competitive industry', *Journal of Political Economy*, **102**, 322–47.
Klepper, S. (1996), 'Entry, exit, growth and innovation over the product life cycle', *American Economic Review*, **86**, 562–83.
Klepper, S. (1997), 'Industry life cycles', *Industrial and Corporate Change*, **6**(1), 145–81.
Klepper, S. (2002), 'The evolution of the US automobile industry and Detroit as its capital', communication at the International Schumpeter Society, Gainsville, FL.
Klepper, S. and Graddy, E. (1990), 'The evolution of new industries and the determinants of market structure', *Rand Journal of Economics*, **21**, 27–44.
Klepper, S. and Miller, J. (1995), 'Entry, exit, shakeouts in the United States in new manufactured products', *International Journal of Industrial Organization*, **13**, 567–91.
Klepper, S. and Simons, K. (1997), 'Technological extinctions of industrial firms: an inquiry into their nature and their causes', *Industrial and Corporate Change,* **6**, 379–460.
Klepper, S. and Simons, K. (1999), 'Industry shakeouts and technological change', mimeo.
Krafft, J. (ed.) (2000), *The Process of Competition*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Krafft, J. (2002), 'Industrial dynamics', in W. Lazonick (ed.), *IEBM Handbook of economics*, London: Thompson Learning.
Lamoreaux, N., Raff, D. and Temin, P. (eds) (1998), *Learning by Doing in Firms, Markets and Nations*, Chicago: University of Chicago Press.
Lazonick, W. (1991), *Business Organization and the Myth of the Economy*, Cambridge: Cambridge University Press.
Malerba, F. and Orsenigo, L. (1993), 'Technological regimes and firm behavior', *Industrial and Corporate Change*, **2**, 45–72.
Malerba, F. and Orsenigo, L. (1996), 'The dynamics of evolution of industries', *Industrial and Corporate Change,* **5**, 51–87.
Malerba, F., Nelson, R., Orsenigo, L. and Winter, S. (1999), '"History-friendly" models of industry evolution: the computer industry', *Industrial and Corporate Change*, **8**, 3–40.
Marshal, A. (1890), 'Some aspects of competition', reprinted in A.C. Pigou (ed.), *Memorials of Alfred Marshall*, London: Macmillan.
MarshalL, A. (1920), *Principles of Economics*, London: Macmillan.
McKelvey, M. (1996), *Evolutionary Innovations: The Business of Biotechnology*, Oxford and New York: Oxford University Press.
Mitchell, W. (1995), 'Medical diagnostic imaging manufacturers', in G. Caroll and M. Hannan (eds), *Organizations in industry*, Oxford: Oxford University Press.
Mueller, D. (1990), *The Dynamics of Company Profits*, Cambridge: Cambridge University Press.
Mueller, D. (1997), 'First-mover advantages and path-dependence', *International Journal of Industrial Organization*, **15**, 827–50.
Nelson, R. (1998), 'The co-evolution of technology, industrial structure and supporting institutions', in G. Dosi, D. Teece and J. Chytry (eds), *Technology, organisation and competitiveness: perspectives on industrial and corporate change*, Oxford and New York: Oxford University Press.

Nelson, R. and Winter, S. (1982), *An Evolutionary Theory and Economic Change*, Cambridge, MA: Harvard University Press.
Pavitt, K. (1984), 'Sectoral patterns of technological change: towards a taxonomy and a theory', *Research Policy*, **13**, 343–74.
Saviotti, P. (1998), 'On the dynamics of appropriability or tacit and of codified knowledge', *Research Policy*, **26**, 843–56.
Schumpeter, J. (1912), *The Theory of Economic Development*, Cambridge, MA: Harvard University Press.
Schumpeter, J. (1942), *Capitalism, Socialism and Democracy*, New York: Harper.
Utterback, J. and Suarez, F. (1993), 'Innovation, competition and industry structure', *Research Policy*, **22**, 1–21.
Van Dijk, M. (1998), 'Industry life cycles in Dutch manufacturing', Working Paper, Center for Research of Economic Micro-Data (CeReM) of Statistics Netherlands.
Windrum, P. and Birchenhall, C. (1998), 'Is product life cycle a special case? Dominant design and the emergence of market niches through co-evolutionary learning', *Structural Change and Economic Dynamics*, 9, 109–34.

9. High growth with 'old' industries? The Austrian Paradox revisited*

Michael Peneder

1. MOTIVATION AND OUTLINE

Ever since a series of broad investigations into Austria's economic performance at the end of the 1980s (Aiginger 1987), comparisons with other developed countries have revealed severe and persistent deficits in industrial structure, as measured, for example, by the shares of technologically progressive industries in total exports or production. In combination with consistent findings regarding comparatively low levels of R&D and venture capital as inputs to the innovation process, or patents as one specific form of technological output, the notion of an Austrian 'technology gap' (Hutschenreiter and Peneder 1997) became firmly established. However, this negative assessment has been sharply contrasted by Austria's successful macroeconomic development, which has more or less endured throughout the past decades and has been characterized by high levels of labour productivity and income, above average (or at least average) GDP growth and below average unemployment rates. This ambiguity was subsequently coined the 'Austrian Paradox' of 'old' industrial structures but high aggregate performance (Peneder 1999, Tichy 2000).

The Austrian Paradox has raised much policy debate, which, however, has remained highly inconclusive. The major obstacle to a more successful discussion was the unsettled dispute regarding the general relevance of industrial structures to macroeconomic development. If the mainstream of economic theory does not predict any relationship between industrial structure and economic growth, and, in addition, the particular Austrian experience appears to be a clear case against it, why should we bother with structural reforms as long as macroeconomic growth continues to do well? The Austrian Paradox thus became a showcase, exhibiting how the lack of adequate theory and solid empirical investigations can block the development of a coherent concept of economic policy.

This chapter attempts to provide a comprehensive interpretation of the Austrian Paradox, first offering an up-to-date empirical assessment, and second a tentative explanation of the specific socio-political factors involved in the traditional Austrian growth regime. As will be seen, this involves some subtle twists and turns in reasoning. It will pay off in the end by paving the way for a fresh look at industrial and structural policies.

The analysis is structured by the following four questions: (i) How wide really is Austria's technology gap and how good is its growth performance? (ii) Other things being equal, is there any general empirical relationship between industrial structure and macroeconomic growth? (iii) If so, what differences in other factors can explain Austria's satisfactory macroeconomic performance? (iv) Finally, which general functions must an economic system achieve in order to enable structural change *and* aggregate growth?

The following sections take up these questions consecutively. Section 2 describes the essential facts about Austrian economic performance. In Section 3 we briefly summarize the results from a complementary econometric study, which strongly confirmed a significant relationship between meso-structure and macroeconomic performance. Section 4 then presents some characteristics of the specific institutional framework in Austria, which might explain the success of its traditional growth regime. Finally, turning to the future challenges ahead, Section 5 applies ideas from evolutionary economics to propose a dynamic conceptual framework for industrial and structural policy.

2. EMPIRICAL EVIDENCE ON ECONOMIC PERFORMANCE

2.1 Deficits in Industrial Structure

One essential characteristic of the Austrian Paradox is the pronounced lack of specialization in particularly technology-driven industries. But before comparing specialization patterns with the EU, we need to know how industries are classified, and whether these classifications actually have the power to discriminate with respect to certain dimensions of industrial performance. Table 9.1 presents three taxonomies of the manufacturing industry that are used in the subsequent analysis. They were created in a series of research projects undertaken on behalf of the European Commission with the explicit intention to facilitate enquiries into industrial performance with respect to the intangible sources of competitive advantage.[1]

Table 9.2 shows that the taxonomies reveal some interesting patterns with respect to the most widely used indicators in comparative studies on industrial development. Beginning with growth performance, between 1990

Table 9.1 Three taxonomies of the manufacturing sector

		Taxonomy I (Factor inputs)		
Mainstream manufacturing (MM)	Labour-intensive industries (LI)	Capital-intensive industries (CI)	Marketing-driven industries (MDI)	Technology-driven industries (TDI)
		Taxonomy II (Human resources)		
Industries with particularly high shares of . . .	Low-skilled labour (LS)	Medium-skilled, 'blue-collar' labour (MBC)	Medium-skilled, 'white-collar' labour (MWC)	High-skilled labour (HS)
		Taxonomy III (External service inputs)		
Industries with high inputs from . . .	Knowledge-based services (IKBS)	Retail & marketing services (IR&M)	Transport services (ITR)	Other

Source: Peneder (2001a).

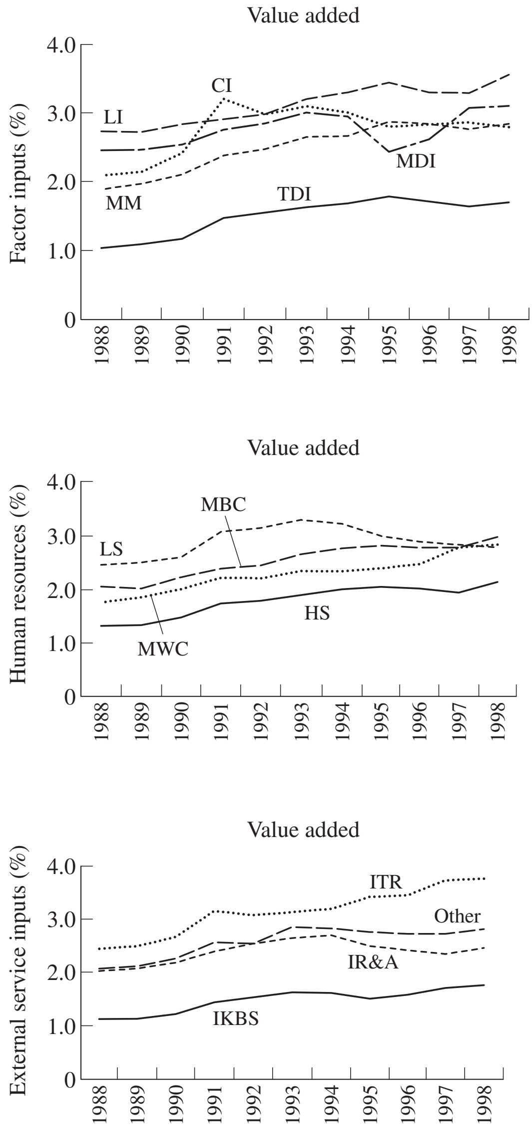
and 1998, annual growth in demand[2] was highest and above 3.0 per cent in technology-driven industries and mainstream manufacturing, high-skill industries, and those branches of manufacturing where external inputs from retail and advertising or from knowledge-based services were exceptionally high. Except for the better performance of marketing-driven industries, a similar pattern emerges if we take a look at value added growth between 1985 and 1998. With regard to employment, the general dynamics depend on the simultaneous interplay of growth in both value added and labour productivity.[3] Finally, unit values in foreign trade are often used to characterize the degree of quality differentiation in various product groups. The numbers reveal that almost all the variation in unit values can be attributed to the separate groups of technology-driven and high-skill industries. Taken together, all the above findings imply that our characterization of industrial structure is highly relevant in terms of industrial performance.

Applying the same taxonomies as Table 9.1, the graphs in Figure 9.1 depict for each type of industry the development of Austria's relative shares in EU value added. It is immediately apparent that within each classification, the taxonomic group which would generally be considered the most

Table 9.2 Industrial specialization and economic performance: EU, Japan and USA

	Consumption (1990/1998)	Value added	Employ-ment	Labour productivity		Export unit values	Import unit values
	EU+Japan+USA					EU	
	Average annual change in 1985/1998 in%				1998 1.000 Euro	1999 Euro/kg	
Mainstream manufacturing (MM)	+3.22	+3.71	−0.17	+3.89	62.1	3.87	3.36
Labour-intensive industries (LI)	+2.50	+3.28	−0.85	+4.17	46.5	2.65	2.41
Capital-intensive industries (CI)	+1.48	+3.54	−1.49	+5.11	102.0	0.56	0.55
Marketing-driven industries (MDI)	+2.47	+4.22	+0.08	+4.14	72.0	1.52	1.33
Technology-driven industries (TDI)	+3.75	+4.74	−0.95	+5.74	104.6	16.26	15.78
Industries with high shares of . . .							
Low-skilled labour (LS)	+2.16	+3.32	−1.01	+4.38	61.4	1.13	1.10
Medium-skilled 'blue-collar' labour (MBC)	+2.92	+4.38	+0.01	+4.37	60.4	4.10	3.41
Medium-skilled 'white-collar' labour (MWC)	+2.83	+4.37	−0.48	+4.88	89.9	1.24	1.11
High-skilled labour (HS)	+3.22	+3.93	−0.58	+4.54	85.5	19.43	17.19
Industries with large inputs from . . .							
Knowledge-based services (IKBS)	+3.07	+4.55	−0.53	+5.11	91.5	2.63	2.05
Retail & advertising (IR&A)	+3.24	+4.53	−0.16	+4.70	79.1	4.62	4.49
Transport services (ITR)	+2.37	+3.71	−0.31	+4.03	67.2	0.86	0.80
Other Industries	+2.24	+3.22	−1.08	+4.35	60.5	1.90	1.73
Total manufacturing	+2.69	+3.94	−0.56	+4.53	72.8	1.91	1.71

Sources: EUROSTAT, WIFO.



Note:
For abbreviations, see Table 9.2.

Sources: EUROSTAT, WIFO.

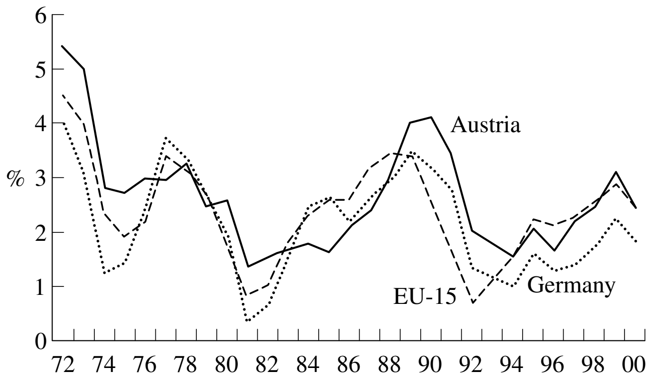
Figure 9.1 Austrian shares of value added and exports in the EU, 1988–98 progressive always exhibits for Austria the lowest value added relative to the European Union. Compared to a share of 2.8 per cent for total manufacturing in 1998, the respective shares are much lower for technology-driven (1.7 per cent) and high-skill industries (2.2 per cent), as well as for industries with high inputs of knowledge-based services (1.8 per cent). There is no evidence of a substantial catching up in terms of industrial structure, as the relative gaps between the various types of industries have not changed in favour of the above categories.[4]

However, despite these persistent deficits in industrial structure, Austria's manufacturing sector increased its aggregate shares of nominal value added in the European Union between 1988 and 1998 by more than one-third. This fact should remind us that industrial structure itself does not constitute a normative goal. It must be derived from the general purpose of raising and maintaining a society's desired standard of living. In other words, these empirical findings are only relevant to economic policy, if industrial structure relates to an economy's level of income and potential for growth. The following section on the performance of the total economy shows that, if taken on its own, the case of Austria could be raised as an illustrative example against such a link.

2.2 Macroeconomic Success

The good macroeconomic performance of the Austrian economy during the last decades constitutes the other essential characteristic of the Paradox. In international comparisons, the variable cited most often in this context is Austria's favourable labour market performance. At a level amounting to 3.7 per cent of the total labour force in the year 2000, the unemployment rate in Austria is well below the European average of 8.2 per cent. The general standard of living is conceived as being relatively high: in 1999, Austria's GDP per capita in purchasing power parities amounted to \$24 646.[5] Average GDP per capita for the EU was \$22 433. Comparing long-run macroeconomic performance between 1970 and 2000, the average growth of GDP per capita at constant prices was 2.7 per cent per annum, surpassing not only Germany (2.2 per cent) but also the European Union (2.4 per cent).[6]

A closer examination of the macroeconomic performance highlights two remarkable features (Marterbauer 2001). First, investment activities were continuously more dynamic in Austria than in the EU. In Austria, aggregate investment grew on average by 2.8 per cent per annum, compared to 2.0 per cent in the EU and 1.6 per cent in Germany. Taking the high shares of investment in GDP into account, macroeconomic growth rates in Austria would even have to be regarded as quite disappointing (especially



Sources: OECD, WIFO, Marterbauer (2001).

Figure 9.2 Gross domestic product (percentage change over previous year; three-year moving averages)

during the 1990s). The discrepancy, however, can be explained by an unfavourable structure of investment, which is heavily biased towards construction and housing.[7] The consequence is a low capital productivity as reported, for example, in Scarpetta et al. (2000). This finding deserves especial attention since construction is heavily influenced by public policy. In Austria the public sector has been an important investor in infrastructure itself, and private investment – for instance in the construction of dwellings – is still highly subsidized.

Second, Austria's growth performance exhibits a very peculiar dynamic pattern. Figure 9.2 reveals that it is repeatedly and consistently superior to the EU and OECD average only during periods of recession (1975, 1981 and 1993).[8] As it appears, the orientation of policy towards anti-cyclical impacts and the stabilization of expectations has led to a less volatile business cycle in comparison with other industrialized countries. This achievement is even more remarkable once we consider that Austria is a small open economy, which depends heavily on international trade and foreign direct investment and therefore is much exposed to international changes in the business climate. The most plausible macroeconomic explanation for this good performance during recessions is found in the effects of automatic stabilization through Austria's generous welfare system. Additionally, the public sector might have had a positive impact through the timing of infrastructure investment, which was often extended during periods of recession and the years thereafter. (A distinct microeconomic or structural explanation will be suggested later.) Of particular relevance to macroeconomic debates is that successful stabilization in times of economic downturn had no substantive and immediate dampening effect upon growth during the upturns that followed. Contrary to conventional wisdom, macroeconomic stabilization in times of economic crisis thus appears to have had a lasting impact on the levels of GDP per capita in the case of Austria.

These findings bring to the surface some interesting aspects of the Paradox: first, since the initial intention of stabilization policy is to achieve short-term effects, the construction sector is an attractive target for discrete policy intervention. Active stabilization policy thus tends to aggravate problems concerning the structure of investment. But there should be no doubt that low capital productivity comes at a cost (at least in terms of the foregone opportunities of more profitable investment), even if these are not easily detected when the focus is on GDP per capita. Hence macroeconomic performance would appear less bright if the focus were on total factor productivity. Second, it is reasonable to assume that successful macroeconomic stabilization also dampened the need and thereby reduced the incentives for industrial restructuring. This can still have a negative effect on aggregate growth over the longer run.

In order to check the veracity of the latter possibility, we need to shift our focus from the single case of one country to an econometric study on a much larger international scale. The next section briefly summarizes the results of this research.

3. THE MESO–MACRO LINK IN AN INTERNATIONAL PERSPECTIVE

Investigating the general relevance of industrial structure to aggregate growth ('the meso–-macro link'), the first observation is a distracting lack of interest within the conventional branches of economic theory. The traditional neoclassical growth theories offer no clues as to how the dimension of industrial structure can be integrated. They are characterized by an exclusively macroeconomic focus and rely on strong assumptions of homogeneity. As a consequence, there is simply no room in the models for variety in industrial structure. In contrast, the new endogenous growth theories must be credited for integrating increasing returns (for example via learning and R&D investments) into the canon of formalized theory. In some of these models, especially those with an inclination towards the Schumpeterian idea of creative destruction, industrial structure does play a role, although a very rudimentary one. For example, a research sector which generates spillovers to the rest of the economy can be isolated. However, the problem is that these models remain strictly within the realm of steady state analysis, regarding all deviations from the path of constant rate, balanced growth as mere transitional dynamics. Some of the most deserving proponents of endogenous growth theory are very precise about the limitations of steady state analysis: 'The economy is always a scaled-up version of what it was years ago, and no matter how far it has developed already the prospects for future developments are always a scaled-up version of what they were years ago' (Aghion and Howitt 1998, p. 65). The analysis of structural change obviously does not fit within such a framework.

A distinct view of the growth process is developing within the evolutionary paradigm. Most important to the task at hand is that evolutionary theory characterizes the market economy as a process of continuous change and transformation. Its dynamics are driven by innovation; but '[i]nnovation is a matter of differential behaviour and differential behaviour is the basis for structural change' (Metcalfe 1998, p. 37). As the fundamental diversity of micro-behaviour involves dynamics which are much richer than in steady state growth, simple aggregation cannot do away with the fact that the potential paths of development are various and depend on the idiosyncratic characteristics of an economy, including its sectoral composition of production. From the evolutionary perspective, structural change is therefore an inevitable companion of growth and development.

In a recent publication, Plümper and Graff (2001) have demonstrated a positive and significant impact of export specialization in 'high-tech' industries on aggregate growth. Applying a somewhat different conceptual framework, data panel and method of estimation, Peneder (2003) similarly longs for an empirical validation of the meso-structure/macro-performance hypothesis, which postulates that structural change is a significant determinant of macroeconomic development and growth. Dynamic panel econometrics is applied to test for the impact of specific structural variables on aggregate income and growth. The data panel comprises 28 OECD countries; the time frame is 1990 to 1998. The data indicate the levels and growth rates of GDP per capita at PPPs. The set of regressors includes demography, employment rates, capital investment and average years of education. These are complemented by the value added share of the services sector and relative shares in the exports and imports of technology-driven and high-skill industries.

In short, for the general macroeconomic variables, the standard results of cross-country regressions are mostly reproduced. But the additional inclusion of several structural variables generates the following stylized facts:

1. Although the share of the services sector is positively correlated with income levels, its lagged levels have a negative impact on annual growth

of GDP per capita after inclusion of the standard variables of cross-country growth regressions. The effect is small, but significant and robust. It is generally consistent with Baumol's hypothesis of a structural burden through unbalanced growth, caused by an inevitable shift of employment towards those branches of the services sector where gains in labour productivity are harder to achieve.

2. Turning to the manufacturing sector, for both technology-driven and high-skill industries, the shares of total exports relative to the OECD have a pronouncedly positive and consistently significant impact on the level and growth of GDP per capita. This observation might be attributable to the influence of two separate channels: from a capability-based perspective, a higher share of 'entrepreneurial' types of industry also implies a higher capacity for aggregate growth in the economy. This explanation offers a direct link to aggregate developments via differential growth rates favouring industries with a greater capacity to expand the consumers' willingness to pay. An indirect link is provided by positive externalities between industries. In the case of producer-related spillovers, proximity (either spatial or institutional) allows for a better diffusion of relevant knowledge within common territorial boundaries.
3. It is not only an increase in exports but also in imports, and hence the application of technologically advanced products, which contributes positively to aggregate growth. Coefficients are smaller and less robust. But the significant impact of the lagged levels of shares in total imports relative to the OECD confirms that user-related rent spillovers (intrinsic to certain goods but independent of proximity in production) are also relevant. Positive externalities between industries arise when the economic value of embodied knowledge is not entirely captured in the prices for intermediate goods.

The essential message is quite obvious. In large-scale international comparisons, structural change with respect to specific types of industry is a significant determinant of aggregate income levels and growth. The empirical evidence thus substantiates the evolutionary concern with Schumpeterian economic development, which regards growth and structural change as two inseparable elements. Hence this study also confirmed our specific concern with the Austrian Paradox. Other things being equal, industrial structure does matter. Over the long term, the lack of structural change in favour of technologically progressive industries implies the danger of a 'growth penalty'. The immediate consequences appear to be small, and therefore easily escape public attention. But industrial structures are quite persistent; in the long run, seemingly small variations in growth rates can accumulate into considerable differences in income levels.

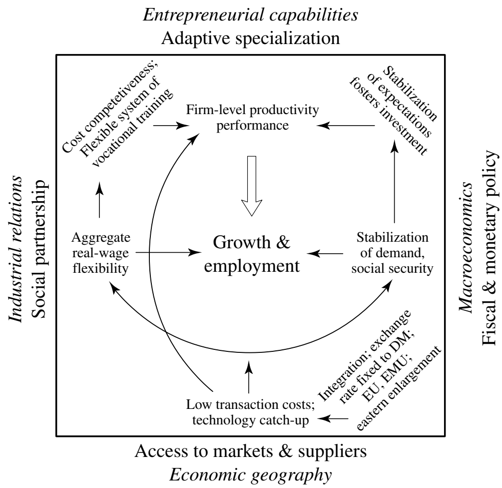


Figure 9.3 Fitting together some pieces of the Austrian growth regime

4. PUTTING TOGETHER SOME PIECES OF THE AUSTRIAN 'GROWTH PUZZLE'

The results of the econometric analysis that were summarized in the previous section are based on the crucial assumption of 'other things being equal'. Its essential message is that the Austrian technology gap must be taken seriously as an indication of a potential growth penalty resulting from persistent deficits in industrial structure. Thus the Paradox is relevant. Nevertheless, the satsfactory growth performance suggests the importance of other determinants, which so far have been able to compensate for Austria's deficits in industrial structure. In order to understand the Paradox and its consequences for economic policy, two critical questions arise: (i) what other things are different and therefore special about the Austrian growth regime?; and (ii) can we expect these elements to be sustained?

The issue is, of course, too broad and complex for any definite answers. In the following brief excursion I can only touch upon some selected aspects of the traditional Austrian growth regime. Four interrelated dimensions of economic performance are reviewed in the schematic representation in Figure 9.3: (i) economic geography, as it determines the potential for market and supplier access; (ii) macroeconomic policy, comprising fiscal and monetary policies; (iii) industrial relations, shaped by the corporate institution of social partnership; and (iv) a particular aspect of entrepreneurial capabilities, summarized under the heading of 'adaptive specialization'. Interactions among these factors are of course manifold. For example, spatial proximity and economic integration can reduce transaction costs and thus enable faster catching up to state-of-the-art technology. Fiscal and monetary policies have an impact on the formation of expectations regarding the business environment, again affecting investment and productivity growth. The nature of industrial relations, which, for example, determine the results of the wage bargaining process, obviously influence real-wage flexibility and cost competitiveness. The combined impact of these and other elements can either be more conducive or obstructive to firm-level productivity performance.

4.1 Economic Geography

It is a well-documented empirical fact that economic interactions decline very rapidly as distance increases (Venables 2001). The New Economic Geography takes this matter seriously by allowing the costs of transaction (costs accrued through searching, quality control, transport, the synchronization of processes, the passage of time and so on) to decrease with spatial proximity. As a consequence, locational choices are explained not only by relative factor prices, but depend also on the costs of market access and the costs of intermediate products, which both tend to increase with distance. Additional complications arise through the presence of positive externalities, for example through pooled markets for skilled labour or the easier diffusion of relevant knowledge, which can trigger the typical dynamics of cluster formation. Otherwise, given equal technology and primary factor prices, central locations which benefit from the cost advantages of spatial proximity can afford to pay higher wages, whereas peripheral regions can only compete with lower labour costs.

The fast pace of theoretical developments in this field has recently come to bear fruit in a series of new empirical studies. According to these, the forces of economic geography appear to favour the Austrian economy. According to data for market potential, which are based on the inverse relationship of incomes to distance for regional disaggregations, Austria is among the countries with the highest values, together with Benelux, France, Germany and the United Kingdom (Midelfart-Knarvik et al. 2000, p. 52). It is only 7.8 per cent below those of Belgium, which is the most centrally located European economy, and 37.6 per cent above the median. Similar calculations can be done for market and supplier access, defined as the distance-weighted sum of the market/supplier capacities of all trading partners. Econometric analysis demonstrates that the two measures of economic geography can explain a substantial part of the variations in per capita income. Due to its proximity to the dynamic, high-income regions of southern Germany and northern Italy, Austria is placed among the countries with particular high levels of access to markets and suppliers (Redding and Venables 2001).

These results are also consistent with an interrogation of top managers of national and international enterprises located in Austria, who placed especial emphasis on the benefits of 'double integration', that is, EU membership plus strong commercial ties to Central and Eastern European countries (Aiginger and Peneder 1997). This source of competitive advantage has not only been sustained, but has grown considerably during the 1990s, due to the dynamics of economic transition in our nearby neighbourhood. The accession of these countries to the European Union can only strengthen Austria's competitive advantage further, by making Austria an even more central location in which to conduct international business. Policy thus has a positive role to play in promoting the eastern enlargement of the EU. Unfortunately, this spatial aspect of economic development is not yet fully recognized in current policy debates about EU eastern enlargement.

4.2 Macroeconomic Policy

Several authors attribute Austria's growth performance since the 1970s to its strong concern for anticyclical stabilization, which was reflected in a coherent concept of demand-oriented macroeconomic policy. Guger (1998) summarized the core elements of this so-called Austro-Keynesianism: first, absolute priority for full employment and growth; second, anticyclical fiscal policy through built-in stabilizers in the generous system of social security, public investment, accelerated depreciation schemes and so on; and third, a strong currency which keeps inflation under control. Together with the fourth element of aggregate real-wage flexibility (Hofer and Weber 2000), this specific policy combination stabilized expectations and thus created a favourable environment for private investment.

We must however mention that the lustre of Austria's macroeconomic performance has not been sustained over the decades. The pronounced positive growth margin of the 1970s largely disappeared and gave way to average performance (Marterbauer 2001). Furthermore, low official numbers on unemployment were increasingly misleading, due to hidden unemployment which developed as a result of early retirement schemes. This degeneration from high to medium levels of performance can be attributed to various aspects related to macroeconomic policy: (i) there is good reason to believe that demand-pull strategies are most effective, the further an economy is operating below its potential productivity level, especially when the catching up through imported technology substitutes for more rigorous structural reforms on the supply side. (ii) Higher growth rates due to catching up also ease the social tensions arising from distributional conflicts. Conversely, when aggregate growth slows down, the public willingness to pay for fiscal expansion decreases and political priorities change in accordance. (iii) Economic integration in general tends to narrow down the potential to pursue divergent national policies; the decision to join the Economic and Monetary Union (EMU) particularly affected Austria's monetary and fiscal policies. National discretion in monetary policy was formally abolished, although the effects of this restraint may have been relatively subtle, since the exchange rate of the Schilling had previously been fixed to that of the German Mark. Discretion in fiscal policy was constrained by an obligation to comply with the convergence criteria (Breuss 2000). In short, national, demand-oriented macroeconomic policies no longer provide a comparable source of high aggregate performance.

4.3 Industrial Relations

Austria's macroeconomic performance cannot be separated from the distinctive role of its corporatist system. The institution of social partnership allowed organized interests to participate in the process of policy formulation, thereby increasing internal consistency in and compliance with the decisions made. Theoretic rationales can be found in the notion of 'social capabilities', emphasizing the importance of ongoing relationships, as well as the creation of consensus and trust, in order to limit transaction costs, uncertainty and conflict (Butschek 1995). Low unemployment, high standards of social security, the much-heralded apprenticeship training system, a high degree of aggregate real-wage flexibility, and the maintenance of a competitive edge in labour cost per unit of output (Guger 2000) all demonstrate how remarkably well the Austrian economy has performed with respect to the core competencies of the social partners. In their comparative study, Casey and Gold (2000) attribute this to an exceptionally high degree of coherence in policy concepts, directed by the shared macroeconomic goals of securing growth and employment:

> In the case of Austria, critical actors were faced with the task of rebuilding a society that had been shattered politically and economically after the war. They therefore constructed a set of procedures and institutions that gave recognition to the interrelationships between taxation and expenditure policy, interest rate and exchange rate policy, industrial assistance and training policy, and social security and health and education policy. . . . Moreover, the act of fitting the parts together, and subsequently of keeping them together, was made possible only because those who were responsible for any one part were simultaneously responsible for other parts too. This meant that *they had the picture as a whole*

> *in their heads*, the nature of their own contribution to it and the ability to recognise the parts that were contiguous and the way in which the matching should occur. (Casey and Gold 2000, p. 29; emphasis added)

In recent years, however, the system has increasingly come under strain. Any institution which is so intricately involved in the policy process is held equally responsible for the less favourable aspects of economic performance. The above authors, for example, argue that corporatist systems cope well with incremental adjustments, but perform badly when major changes in direction are needed. A related problem is an excess of concern for the established interests of insiders, which tends to harm the potential interests of outsiders. Such a system is therefore less dedicated to promoting policies oriented towards future innovation or entry conditions advantageous to new entrepreneurs. In addition to widespread concerns about the democratic legitimacy of the corporatist system, the fact that a consensus can be attained only through complex negotiations has been accused of hindering the timely delivery of necessary reforms. As a consequence, governments have increasingly tried to emancipate themselves and bypass the system of social partnership in the introduction of new legislation.

4.4 Adaptive Specialization

Economic geography, government policy, corporatist and other institutions can only shape the general business environment of locations, making them either more conducive or obstructive to entrepreneurial activity. An important piece, namely the particular nature of entrepreneurial capabilities, is missing from the puzzle and calls for at least a tentative explanation. Despite the apparent danger of oversimplification, a strong case can be made for the entrepreneurial capability of 'adaptive specialization' as a major source of firm-level productivity performance. Several empirical findings support this conclusion.

Unit values, which are a measure of quality and the vertical differentiability of products in foreign trade statistics, indicate that Austrian manufacturing competes successfully in high-quality goods (Aiginger 2000). Export unit values are close to the European average and Austria is ranked sixth with regard to export shares in the segment of industries which characteristically have the highest unit values. At first sight, this might not seem to be a big achievement. But, given Austria's structural deficits, and especially its low share of technology-driven industries (that is, the one group which usually surpasses other industries by far in terms of unit values; see Table 9.2), average aggregate unit values imply that Austrian manufacturing primarily serves the high-quality segments within its more traditional patterns of specialization.

This interpretation is further strengthened by data from the Community Innovation Survey (Leo 1999). Within most industries, innovative performance in Austria corresponds to EU averages or is even better. This implies that the desired increase in total R&D expenditures, currently the major preoccupation of Austria's innovation policy, is foremost a problem of industrial structure. Two additional observations from the CIS data are of relevance here: (i) differentiation by firm size reveals that it is only the small enterprises that surpass their EU counterparts in terms of innovative activities; (ii) in Austria, customer relations are a much more important source of information for innovative activities than in the EU as a whole.

The Austrian Technology Delphi (Tichy 2000), which screened various markets with respect to their potential for leadership of Austrian firms, is another source of supportive evidence. It demonstrates that the patterns of sectoral specialization are deeply rooted in the corresponding pattern of industrial and scientific expert knowledge. Tichy concludes that

> Austria's high share of basic and mainstream industries rests on a solid long-maintained knowledge base in several fields, which is usually regarded as medium technology (e.g. advanced materials, vehicle components, machinery, . . .). The concentration on intermediate rather than consumer goods assigns prime importance of receiver competence towards commercial customers. (Ibid., p. 434).

This interpretation is also supported by the relatively high importance of small to medium-sized enterprises (OECD 2000), many of which have developed their specific competitive advantages as flexible and reliable suppliers to other industrial customers.

Finally, supportive evidence is also found in the comparatively large inflows of foreign direct investment. The many affiliates of multinational enterprises do not just increase their existing stocks of capital investment, but additionally support the transfer of knowledge and firm-specific assets, including the set of established customer relationships. In a recent study on Austrian manufacturing, Egger and Pfaffermayr (2001) used this as an explanation of their econometric finding that in addition to the effects of capital deepening, foreign direct investment also has a separate and significant impact on labour productivity. In short, both characteristics – the importance of small and medium-sized enterprises (SMEs) and large inward foreign direct investment (FDI) – enhance the overall capability to adapt to and participate in the industrial dynamics of foreign markets.

To conclude, the various findings support the view that flexible specialization and adaptation to specific customer needs is a major strength, which explains a considerable part of Austria's high-productivity performance at the micro-level. Innovations may be largely incremental and therefore of comparatively modest size. But they are widely distributed across a large number of actors, fostering the overall capacity to benefit from proximity to some of Europe's most dynamic economic areas. Furthermore, adaptive specialization – a competency largely established within industrial supply relationships instead of final consumer goods – also offers an additional structural explanation for Austria's superior performance during periods of international downturns (demonstrated in Figure 9.2).

5. PARADIGM SHIFT: TOWARDS A DYNAMIC CONCEPT OF INDUSTRIAL POLICY

The long-term coherence between macroeconomic policies and the other institutions, which bear responsibility for organized interests, is generally considered to be among the most remarkable features of the traditional Austrian growth regime. This coherence was possible only because the goals of macroeconomic policy and the social partners were united through a minimum number of concepts shared by both. Given the specific historic situation and Austria's geographic location, the resulting policy mix effectively fostered growth and employment, based upon economic integration, the import of advanced technology, demand stabilization, high investment, the flexibility of aggregate real wages, and a system of apprenticeship training, which adapts quickly to the actual demand for labour skills. Due to these factors, the system performed well, even though political responsibilities for enabling innovation and structural change were largely neglected.

Due to the various reasons explained above, the traditional growth regime has lost much of its lustre. Not all of the above elements could be sustained. The eventual loss of Austria's positive growth margin is an unmistakable empirical indication (Marterbauer 2001). The potential for additional growth through technological catching up disappeared as early as the 1980s. Membership in the Economic and Monetary Union marked another turning point, at which time the concept of a national, demand-oriented macroeconomic policy lost much of its feasibility. The low share of dynamic industries is additionally threatening to dampen the prospects for future growth. Given this situation, the required policy mix will inevitably shift towards the supply side, and industrial policy will therefore have to be prepared to generate additional impulses for long-term growth through structural reforms. After drifting already from high to average macroeconomic growth, the Austrian Paradox might otherwise resolve into a situation of old structures and 'low performance' (Tichy 2000).

Raising the share of R&D investment in GDP is currently one of the most important benchmarks of Austrian economic policy. But the empirical evidence has demonstrated that if any substantial increase in overall innovative activity is desired, a considerable increase in the dynamics of structural change will first have to be achieved. The Austrian technology gap is not just a problem of public support for R&D. It is more deeply rooted in the lack of structural change towards those types of industries in which technological opportunities relatively abound, thereby attracting more private investment into innovation.

Trying to support or even trigger the dynamics of structural change is, however, a highly complex matter. To begin with, economic policy must encompass a common vision, shared goals and a coherent concept for the different actors involved. As was the case with the traditional model of demand-oriented macroeconomic policy, microeconomic structural reforms require an equally coherent framework in order to be effective. The conventional workhorse for such a framework can be found in the static welfare analysis of so-called market failures, an approach which must be acknowledged for its well-defined target and clear policy prescriptions. The various rationales of policy intervention are well known and need not be repeated here. A fundamental critique, however, has been ever present from the Austrian School of economics, where the market failure approach was rejected due to its misleading use of perfect competition as a point of reference. As the Austrian School has pointed out, perfect competition is tantamount to the impossibility of entrepreneurial innovation. Furthermore, since perfect competition is the mere theoretical construct of an idealized state, untenable in the actual business world, shrewd intellectual exercise can always establish potential welfare gains through the interventions of a presumably omnipotent and benevolent government (Kirzner 1997). In contrast to its free market rhetorics, the analytic benchmark of perfect competition distracts our attention away from the creative forces of self-organizing processes towards the legitimization of rather crude and mechanical policy interventions.

An alternative to static welfare analysis is offered by the evolutionary perspective. Although methodologically still less elaborate, it offers the clear advantage of congruence with a dynamic vision of industrial development. Peneder (2001a) offers a stylized projection of the fundamental rationales of evolutionary dynamics into a coherent set of economic policies, which are directed towards Schumpeterian development instead of mere (steady-state) growth. It cannot be fully elaborated here. But Figure 9.4 summarizes the major intuition in a simplified and very stylized illustration. Its purpose is to define those general functions which the economic system must achieve in order to enable both growth *and* structural change. The idea rests upon the well-known trinity of fundamental forces of evolutionary change: (i) variation, (ii) cumulation and (iii) selection. It is based on the view that any kind of evolutionary change, including tech-

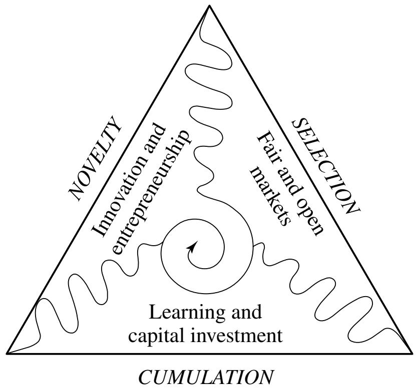


Figure 9.4 Three requirements for an economic system to achieve growth and structural change

nological innovation or structural adjustment, depends on the simultaneous interplay of all three forces. With slight variations, these Darwinian principles regularly appear in the major texts on evolutionary economics.[9] Hodgson (2001) demonstrates why these principles can be regarded as a kind of meta-theory, which is not confined to the realm of biology, but rather operates at a higher level of abstraction.[10]

These very abstract principles are then projected into the realm of economics, defining what the system should achieve in order to support Schumpeterian industrial development: (i) the introduction of novelty through innovation and entrepreneurship; (ii) the accumulation of productive resources through learning and capital investment; and (iii) competitive selection through fair markets. Appropriate policies can then be assigned to these specific functions. Examples are R&D policies or start-up policies, each of them fostering the entrepreneurial function of generating novelty. Education and training, as well as investment policies, including the marketing of business locations to foreign investors, can be attributed to the second pillar of learning and capital accumulation. Finally, competition policy, market liberalization and economic integration are prime examples of policies responsible for guarding fair principles of selection in open markets.

Such a general scheme is of course only a first, and indeed very modest, step towards a better understanding of the dynamic rationales for structural policies and their mutual dependence. The intention is not to invent any new policies, but to present relevant fields of structural policies within a coherent framework. This framework should offer some flexibility, but nevertheless be derived from fundamental theoretical considerations about industrial dynamics and structural change. What makes an important difference to the traditional arguments based upon static welfare analysis is the compelling need to legitimate and design the various policies interdependently, that is, with regard to the overall functions of the economic system they are meant to support. In contrast to the static welfare approach, it is no longer sufficient to justify any policies according to their own particular rationales (the presence of externalities, information asymmetries, public goods, subadditivity of cost and so on) and execute them in isolation from one another. It is the systemic nature of the evolutionary perspective which forces upon policy a shared vision of the overall goals and the specific tasks for which each of its disciplines is responsible. Although Figure 9.4 is only an exemplary exposition, it might serve well as an illustration of the paradigmatic shift required to achieve a dynamic process of 'Schumpeterian' development as a combination of high growth *and* structural change.

6. SUMMARY AND CONCLUSION

This chapter was motivated by the repeated observation of pronounced deficits in the industrial structure of the Austrian economy – strikingly contrasted by the general perception of its good macroeconomic performance. Econometric studies confirmed that structural change with respect to specific types of industry is a significant determinant of aggregate income levels and growth. The good growth performance during the past can therefore not eliminate the relevance of structural deficits. It can only suggest the importance of other determinants, which in the case of Austria were more favourable to economic growth. Among these, geographical proximity to dynamic markets and suppliers, a coherent macroeconomic policy, industrial relations and a certain entrepreneurial competence in adaptive specialization have been singled out as important elements. However, not all of the above elements can be sustained. The potential for additional growth through technological catching up vanished as early as the 1980s. Participation in the Economic and Monetary Union marked another turning point, at which the concept of national, demand-oriented macroeconomic policy lost much of its feasibility. The corporatist system is on the retreat, giving way to increasing political pressure. Although the precise impact of all these developments is difficult to assess, it is clear that the characteristic coherence of the traditional Austrian growth regime is breaking down.

In the light of persistent structural deficits in the sectoral composition of Austria's production, a far-sighted concept of economic policy should target Schumpeterian development, characterized by the simultaneous interplay of growth and structural change. The chapter therefore concludes with an emphasis on the three elementary forces which an economic system must deliver in order to foster evolutionary dynamics: (i) novelty through entrepreneurship and innovation, (ii) the accumulation of productive resources through learning and capital investment, and finally (iii) open and competitive markets, which is the economist's standard solution to provide a fair and effective means of selection.

NOTES

* This chapter draws upon the knowledge and experience of many experts on the Austrian economy. My principal thanks go to Markus Marterbauer, Karl Aiginger and Gernot Hutschenreiter, whose original contributions to the topic are independently available in an edited volume (Peneder, 2001b). Important additional advice, comments and suggestions were provided by Heinz Hollenstein, Michael Pfaffermayr, Wolfgang Pollan, Gunter Tichy and Ewald Walterskirchen. Especial thanks are also due to Andrew Wycoff, who joined us for fruitful discussions related to the OECD activities concerning economic growth. Together with her colleagues at WIFO, Eva Sokoll managed all of the necessary supportive data services.

1. Detailed information on the classifications, data sources and methodology is presented in Peneder (2002).
2. The indicator is 'apparent consumption' (comprising industrial production + imports − exports).
3. While we generally observe that manufacturing industries are not able to maintain employment levels, the figures in Table 9.2 make apparent the different sources of decreases in the number of jobs. Some branches, such as low-skill and labour-intensive industries, are primarily hit by the lack of dynamics from the demand side. In others, employment is lost primarily due to their extraordinarily good performance in terms of labour productivity. Technology-driven industries are the best example. But when demand dynamics are low and labour productivity growth is simultaneously high, the combined effect on employment is most detrimental. This corresponds to the actual situation of capital-intensive industries.
4. Viewed from a different angle, in the EU 22.9 per cent of total value added in the manufacturing sector in 1998 was generated by technology-driven industries. In Austria it was only 14.2 per cent. Among EU member states, only Greece (6.7 per cent) and Portugal (10.5 per cent) have lower value added shares of technology-driven industries, whereas Sweden (29.8 per cent) and Ireland (29.6) have the highest. France (27.2 per cent), Germany (26.4 per cent) and the UK (24.3 per cent) are also above the EU average. All the other countries have value added shares in the range between 15 per cent and 20 per cent. See also Peneder (2000).
5. This figure is only marginally lower than those for Belgium and the Netherlands, which are ranked fourth and fifth within the European Union.
6. During the same period, employment grew on average by +0.4 per cent per annum in Austria, corresponding to the same number for the EU and surpassing Germany (+0.3 per cent). At nearly 70 per cent, the current employment rate is surpassed only by Denmark, Sweden, the Netherlands and the UK (while the European average amounts to only 62 per cent).
7. At 24 per cent of GDP, Austria's investment ratio in real terms was much higher than the European average (20 per cent of GDP). However, construction investment

accounted for approximately 56 per cent of total investment in real terms and around 14 per cent of GDP in 1998, while the corresponding figures are only 47 per cent and 10 per cent for the EU average (for Germany: 57 per cent and 13 per cent).

8. The good performance around the year 1990 can largely be attributed to the additional impetus provided at the beginning of economic transition of the Central and Eastern European countries.
9. To review only the most recent examples, see Hodgson (1993, p. 46), Nelson (1995, p. 54), or Metcalfe (1998, p. 22).
10. Peneder (2001a) also investigates the general validity and applicability of this trinity in the economic domain with respect to the market process (Chapter 1), an entrepreneurial view of the firm (Chapter 2), and the above rationale for structural policies (Chapter 6). None of them relies on any biological interpretation.

REFERENCES

Aghion, P. and Howitt, P. (1998), *Endogenous Growth Theory*, Cambridge, MA: MIT Press.

Aiginger, K. (ed.) (1987), *Die internationale Wettbewerbsfähigkeit Österreichs*, Volumes 1, 2 and 3, Vienna: WIFO.

Aiginger, K. (2000), 'Europe's Position in Quality Competition', European Commission, DG Enterprise, Working Paper.

Aiginger, K. and Peneder, M. (1997), *Qualität und Defizite des Industriestandorts Österreich*, Vienna: WIFO.

Breuss, F. (2000), 'An Evaluation of the Economic Effects of Austria's EU Membership', *Austrian Economic Quarterly*, **5**, 171–96.

Butschek, F. (1995), 'Sozialpartnerschaft aus der Sicht der Neuen Institutionenökonomie', *WIFO-Monatsberichte*, **68**, 644–54.

Casey, B. and Gold, M. (2000), *Social Partnership and Economic Performance. The Case of Europe*, Cheltenham UK and Northampton, USA: Edward Elgar.

Egger, P. and Pfaffermayr, M. (2001), 'A Note on Labour Productivity and Foreign Inward Direct Investment', *Applied Economics Letters*, **8**, 229–32.

Guger, A. (1998), 'Economic Policy and Social Democracy: The Austrian Experience', *Oxford Review of Economic Policy*, **14**, 40–59.

Guger, A. (2000), 'Verbesserung der relativen Lohnstückkostenposition durch Euro-Kursrückgang', *WIFO-Monatsberichte*, **73**, 541–6.

Hodgson, G.M. (1993), *Economics and Evolution. Bringing Life Back to Economics*, Cambridge: Polity Press.

Hodgson, G.M. (2001), 'Darwinism in Economics: From Analogy to Ontology', mimeo (10 November).

Hofer, H. and Weber, A. (2000), 'Wage Mobility in Austria 1986–1996', mimeo, Vienna: IAS.

Hutschenreiter, G. and Peneder, M. (1997), 'Austria's "Technology Gap" in Foreign Trade', *Austrian Economic Quarterly*, **2**, 75–86.

Kirzner, I.M. (1997), 'Entrepreneurial Discovery and the Competitive Market Process: An Austrian Approach', *Journal of Economic Literature*, **35**, 60–85.

Leo, H. (1999), *Die Innovationsaktivitäten der Österreichischen Wirtschaft*, Vienna: WIFO.

Marterbauer, M. (2001), 'The Loss of the Positive Growth Margin', in M. Peneder (ed.), *Structural Change and Economic Growth*, Vienna: WIFO, pp. 33–52.

Metcalfe, J.S. (1998), *Evolutionary Economics and Creative Destruction*, London: Routledge.

Midelfart-Knarvik, K.H., Overman, H.G., Redding, S.J. and Venables, A.J. (2000), 'The Location of European Industry', mimeo, London: LSE, p. 52.

Nelson, R.R. (1995), 'Recent Evolutionary Theorizing about Economic Change', *Journal of Economic Literature*, **33**, 48–90.

OECD (2000), *OECD Small and Medium Enterprise Outlook*, Paris: OECD.

Peneder, M. (1999), 'The Austrian Paradox: "Old" Structures but High Performance?', *Austrian Economic Quarterly*, **4**, 239–47.

Peneder, M. (2000), 'Intangible Assets and the Competitiveness of European Industries', in P. Buiges, A. Jacquemin and F. Marchipont (eds), *Competitiveness and the Value of Intangible Assets*, Cheltenham, UK and Northampton, USA: Edward Elgar.

Peneder, M., (2001a), *Entrepreneurial Competition and Industrial Location*, Cheltenham, UK and Northampton, USA: Edward Elgar.

Peneder, M. (ed.) (2001b), *Structural Change and Economic Growth*, Vienna: WIFO.

Peneder, M. (2002), 'Intangible Investment and Human Resources', *Journal of Evolutionary Economics*, **12**, 107–34.

Peneder, M. (2003), 'Industrial Structure and Aggregate Growth', *Structural Change and Economic Dynamics*, **14** (4), 427–48.

Plümper, T. and Graff, M. (2001), 'Export Specialization and Economic Growth', *Review of International Political Economy*, **8**, 661–88.

Redding, S.J. and Venables, A.J. (2001), 'Economic Geography and International Inequality', mimeo, London: LSE.

Scarpetta, S., Bassanini, A., Pilat, D. and Schreyer, P. (2000), 'Economic Growth in the OECD Area: Recent Trends at the Aggregate and Sectoral Level', OECD Economics Department Working Paper.

Tichy, G. (2000), 'The Innovation Potential and Thematic Leadership of Austrian Industries. An Interpretation of the Technology Delphi with Regard to the Old Structures/High-Performance Paradox', *Empirica*, **27**, 411–36.

Venables, A.J. (2001), 'Geography and International Inequalities: The Impact of New Technologies', London: LSE (paper presented at the Annual Conference of the Austrian Economic Association, Graz, 17–18 May).

10. Changing structure – keeping location: from musical instruments to electronics in the accordion district of Ancona

Deborah Tappi*

The central idea of economics, even when its foundation alone are under discussion, must be that of living force and movement (Marshall 1920)

1. INTRODUCTION

This chapter aims to provide an empirical contribution to the question of how qualitative change occurs within specialized, local industrialized clusters (LICs). A descriptive model is used to explain the development of Italian LICs by testing certain assumptions in a case study on the musical instrument cluster of Ancona (Marches, Italy). In particular, the forces and processes that allowed this LIC to transform its production structure from very traditional artisanry to innovative and technology-intensive production is investigated. The descriptive model uses a framework in which exogenous changes and internal learning processes trigger a transformation in the production structure. The concepts of structural change, self-organization and learning will be employed in order to provide a conceptual framework to understand the evolutionary trajectories of the cluster over the last decades. The basic aim is to provide a framework for the understanding and explanation of the transformation of LICs.

Of particular interest is the interaction between demand and the process of internal selection of new technologies and new products. Taken together, these lead to a change in the production structure of the LIC. The process of selection of new products to be integrated in the production process and of new markets to be explored depends on the technological path and on complementarities between new information and the accumulated available knowledge, skills and competencies in the LIC. Incumbent firms explore possible alternative supply patterns by trying to apply their skills, competence and knowledge stocks to new production niches. The case study focuses on the internal structural transformation of the musical instrument cluster in Ancona from the 1950s onwards. The case shows that the capacity to absorb information and to adapt to exogenous change was the guiding force behind the self-organizational processes which allowed the LICs to evolve.

There are many theories which address changes in economic systems by emphasizing self-organization processes and endogenous forces (Dosi 1984, Dosi et al. 1988, Foss 1997, Foster 1997, Foster and Wild 1999, Metcalfe 1998, Witt 1997 and 1998). However, this literature has not applied to the study of LICs. Although behavioral geography applies some ideas common to evolutionary thought to the study of agglomeration economies, an evolutionary perspective in explaining the reasons which determine the success of LICs has not been widely explored. The debate on Italian industrial districts has focused attention on their definition and has not tried to answer the question of which factors determined their evolution. However, there is a need to focus on the issue of transformation over time and to point out the forces that shaped the direction of this process. Some works aim to introduce an evolutionary perspective into the study of LICs. Boschma and Lambooy (1999) applied key concepts of evolutionary economics to key issues of economic geography in order to address the question of why regions have different abilities to apply new varieties, and what economic and institutional structures determine a region's competitive advantage. Fujita (1996), in a review of recent economic geography literature discusses the relationship between spatial agglomeration and economic growth. Despite the frequency in which the words evolution and path-dependence appear in the literature on LICs and industrial districts, only recently has an evolutionary perspective begun to be explored (Antonelli 2000, Belussi and Pilotti 2000, Brenner 2000, Lombardi 2001, Orillard 1997).

The present chapter first provides a conceptual outline of how specialized local skills and knowledge interact with new demand patterns. The conceptual outline is explored by means of a case study. The chapter shows that the process of change in the production structure in the LIC of Ancona from musical instruments to electronics was achieved by adding new products to the product portfolio of the LIC. This led to a transformation of the relational and the technological structure of the LIC, and in turn to more differentiated local production networks that are connected, but different from the original.

The chapter proceeds as follows. Section 2 presents an stylized conceptual framework for the common elements of the evolution experienced by many Italian LICs. The factors responsible for the emergence and evolution of

Italian LICs are identified. Section 3 describes the role played by specialization patterns within the LIC of Ancona. Section 4 reconstructs the phases which determined structural changes in the district on which the analysis is based, from the 1950s through the different transition phases to the present structure. The regional economy experienced a peculiar evolutionary pattern, taking place through progressive adjustments and radical structural transformations. A short historical account of the origin and evolution is provided. Section 5 provides a discussion and an interpretation of the history in light of the conceptual framework outlined in Section 2 and concludes the chapter.

2. CHANGE IN LICS

Following Porter's definition of clusters, an LIC is identified as 'a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by communalities and complementarities' (Porter 1998, p. 199). In the LICs the final producer, assembling, service firms and supplier interact with local infrastructures and local credit supply to form this particular field of localized complementarities. Of central importance are the supporting institutions providing education, information, specialized training and technical expertise. LICs are often constituted by firms operating in both traditional and high-tech industries. For this reason, it is difficult to derive a picture of these linkages on the basis of standard industrial classifications (Tappi 2002).

The LIC is also a possible locus of overlapping and spatially concentrated interaction of different networks, whose scope can exceed the geographical location of the boundaries of the LIC. In general terms, the notion of a network employed here is that of a changing relational structure framing both production and knowledge interactions (links) among actors located in nodes. Indeed, in many cases it is possible empirically to identify (Lissoni and Pagani 2002) several subsystems or networks, characterized by a higher internal frequency of interaction within the same LIC. The structure of these relationships is subject to change. The endogenous processes that change the interaction structure are nourished by changes in the external environment. The structure we observe is the result of internal adaptive/reactive forces to external events. The evolution of the LIC therefore ensues from exogenous changes and is led by endogenous forces changing the relational and production structure. If knowledge is exchanged and diffused through the network's links, a change in the frequency of interrelation between network actors or a change in the relational structure, such as adding or removing nodes, has an influence on the knowledge and information to which an actor has access.

Qualitative change is the emergence of new activities which are 'qualitatively distinguishable from pre-existing ones, and the disappearance of pre-existing ones' (Saviotti 2001, p. 197). Qualitative change does not come from heaven but requires absorptive and adaptive capacities. Absorptive capacity is the capacity to absorb information in the external environment and knowledge of other techniques and products which can be integrated into the former, while adaptive capacity is the capability to adapt to external changes by altering the internal structure. The concept of absorptive capacity was introduced by Cohen and Levinthal (1989, 1990) concerning firms' capacity to assimilate and exploit existing information. Absorptive and adaptive capacity depend on the kind and amount of knowledge they are exposed to and have access to. The set of solutions to a problem generated by a change in the environment in terms of techniques selected among several depends: (1) on the population of possible solution known and (2) on the previous choices, that is, the degree of complementarity with accumulated technological knowledge.

Qualitative changes of the production structure of LICs are required in the development of an LIC over time. The Italian experience shows this in a very clear way. In those LICs which were able to resist crisis change took place primarily through changes in the production structure. However, the ability for structural qualitative change depends on the specific absorptive and adaptive capacities of the LIC, which in turn are determined by the network structure within the LIC. Section 4 explains why and to what extent accumulated technological knowledge might affect the choices in an LIC. The learning process through which information and knowledge are selected depends on the stock of accumulated knowledge and capabilities within the LIC and the technological trajectory. This introduces a strong form of path-dependence into the development over time. Insights from the theory of the industry life cycle can provide an understanding of the mechanisms of structural change.

The life cycle of LICs can be divided into four phases: (1) the trigger phase (the origin), which very often goes back to the nineteenth century; (2) the incubation phase (industrial scale production); (3) the advanced phase, from the 1980s onwards; and (4) the death (in some cases). This chapter focuses on the second and third phases of LICs' evolution. The transition from the second to the third phase is characterized by a common trend in firms to diversify their product and customer portfolio and by the start-up of new firms in locally new industries. Indeed, the advantage of industry specialization can be related to the phase of the life cycle the LIC is going through. The second phase, which can be identified with what in the literature has been defined as a typology of LICs, namely the 'Marshallian industrial district', is often characterized by competitive advantage of deep specialization and division of labor. The third phase is characterized by faster innovation processes and technological shifts and product diversification. At this stage, not only demand saturation and congestion can determine the LICs decline, but reluctance of firms to switch to new industries and markets (Guerrieri et al. 2001). The decline of many traditional LICs can indeed be attributed to the inability to perceive and explore new opportunities outside the established industry.

For many Italian LICs the transition period from the second to the third phase can be identified as starting in the 1980s, when competition became increasingly international and the diffusion of knowledge accelerated (Lombardi 2001, Lazerson and Lorenzoni 1999). The increasing internationalization and technological change led to changes in global demand and competition which undermined the competitive position of traditional LICs. International demand changed towards products with 'new technological mixes'. International competitors with competitive high-tech products entered the specialized market niches of many LICs. Three main changes in the external environment can be identified as affecting the innovative attitude of LICs: (1) change in demand and taste, (2) external competition, and (3) rapid technological advancement in fields outside the traditional specialization of the LIC.

The changes in the external competitive environment triggered processes of transformation in traditional LICs. Technology is not exogenous and not all firms are able to adapt quickly to changes in demand and external competition. Even the capability to perceive signals from outside the LIC is the same among local actors. Many traditional LICs have undergone a transition process in which structural change in response to exogenous stimuli led to a restructuring of internal networks. Lombardi (2001) emphasizes that LICs 'altered the division of labor and the pre-existent specialization, by triggering a transition process during which LICs have undertaken strategies of response and structural adjustment as regards exogenous shocks'. Change emerged 'due to a stochastic shock which leads the LIC to an adjustment to new environmental conditions'. While we are aware that change may also ensue in the absence of such a threat, this adaptive strategy accounts for the particular case in which an LIC needs to reconfigure new production opportunities because exogenous changes affect negatively the competitive performance of local firms. In this respect we focus on the internal generation of change, not on the exogenous mechanisms (quoted above) which trigger strategies of change.

Most LICs were able to operate qualitative changes leading to different activities or a different organization of activities, that is, new production networks. The traditional local skills and competencies developed were adapted by interaction within the LIC which enabled absorption of external knowledge. The dynamic reconfiguration of an LIC is the result of the co-evolution of several internal and external elements, including exogenous shocks and self-organizational processes. The connecting link between exogenous shocks and internal self-organizational processes are the 'extroverted actors', who developed a broader cognitive frame by participating in non-only-local networks and triggered the local processes of change. The inter-firm division of labor has repercussions on the inter-firm division of knowledge (Belussi and Pilotti 2000). Heterogeneous firms in terms of size, performances, kind of knowledge and information processed populate LICs (Rabellotti and Schmitz 1999). Even though in an LIC one can observe a collective capacity to evolve and survive, it is clear that collective efficiency is the outcome of an internal process in which some enterprises grow and others decline. Lazerson and Lorenzoni (1999) emphasized the role of larger firms, which 'often orchestrate subcontracting relations, explore commercial avenues, and invest in R&D. These diverse activities link leading firms to distant and local actors, putting them in a strategic position to respond quickly to external market demand, while realigning the productive resources of less sensitively located actors'.

Some common processes of internal adaptation from the production side can be recognized across the evolution of different LICs. They are linked to exogenous changes. The internal process of reshaping the organization of production within Italian LICs largely took place by increasing the degree of product differentiation and new technologies. More markets were served, different techniques applied and generally the size of firms changed. The increase in the import of components, knowledge and information from outside the LIC received particular attention in the literature. In general a shift from a local to a global strategy can be observed.

The questions we try to answer in this chapter are: How do traditional LICs change and adapt their structure? What are the key factors responsible for the direction of the change in the productive structure? However, the answers might not be generalizable, as local specificity plays an important role in each LIC. Therefore we concentrate on some common features which can be identified in several LICs.

The next section discusses the case of the evolution of the musical instrument cluster in Ancona (Italy) to highlight the path-dependence of technological and organizational change and the local embeddedness of process in the development of the LIC over time. The emphasis is on the mature phase of the LIC, in which a transition from electronic musical instruments to electronics took place.

3. TECHNOLOGY AND KNOWLEDGE TRANSFER TO AND FROM MUSICAL INSTRUMENT PRODUCTION: THE SEARCH FOR NEW PATHS

The history of musical instrument production provides many examples of technology and knowledge transfer from and to widely separated industries. It is difficult to understand why it received so little attention from researchers. In many cases, transfers involved the mere adoption of existing technologies in (or from) the musical instrument industry, such as the use of plastic instead of paper in the production of cones for loudspeakers. In other cases, the use of new materials involved more effort, such as in the case of cast iron adapted by piano builders to produce a greater volume of sound. A main innovation in musical instrument production was the adoption of the use of valves from the technology of the steam engine. Valves were originally used to control the passage of steam while pumping water out of mines, and adapted in 1815 by Friedrich Blühmel and Heinrich Stölzel to musical instruments. Blühmel, a trumpet player, got to know this technology by working in a mining company, where he studied the use of valves to control the passage of air to blast furnaces and forge ironworks. Stölzel was a horn player, as well as an instrument builder and repairer in Berlin. Blühmel and Stölzel patented a spring-controlled valve mechanism to trumpets and horns so as to be able to produce a continuous musical scale. The electronic valve served as the basis for all audio and television designs until the invention of the transistor in 1947 by three American scientists at Bell Laboratories (who were awarded the Nobel Prize for their invention in 1956). The adoption of transistors spread quickly and is now applied in almost all electronic devices.

The use of valves and the transition to transistors in musical instrument production were key technological innovations and examples of knowledge transfer in the musical instrument industry. As will be shown in the next section, innovations in other industries affected the technological specialization of musical instrument firms and allowed them to explore new markets. The close relation between the instrument industry and other industries made it easy for 'extroverted' firms to find new production niches once the musical industry became unattractive because of international competition. It was the accumulation of technological knowledge that allowed the LIC to first diversify. This gave rise to new networks. It was start-up firms that specialized first in the production of valves and then transistors, in order to supply the local market created by diversifying incumbents searching for new product niches.

The continuous spillovers between the musical industry and other industries show how a locally embedded LIC which specialized only in one product, or in a typology such as the accordion first and then musical instruments, was able to access knowledge on completely different technologies created in different locations. The thesis is that the continuous contact of 'extroverted' local actors with external sources of technological knowledge is instrumental for the whole LIC to organize its transformation.

The importance of the local dimension for a firm's or network's competitive advantage with respect to the embeddedness of knowledge and skills is influenced by (i) access to a specialized labor pool: innovation – especially by established firms and networks – is a path-dependent process (the direction of the change depends very much on the previous technological trajectory and available knowledge and skills which are embodied in the local labor force) and (ii) by the fact that external information and knowledge need to be translated and contextualized by local competence according to local knowledge. An important part of knowledge of the production process, whether tacit or localized, is transferred by means of personal interaction and learning on the job (Johnson et al. 2001).

The evolutionary literature aims to identify the determinants of both industrial dynamics and technological trajectories in the economy (Nelson and Winter 1982, Dosi 1984, Saviotti 2001). Some authors aimed at combining evolutionary thinking with economic geography (Boschma and Lambooy 1999). Studies of the determinants and possible direction of innovation in LICs have been made (for example Breschi 2000, Camagni 1991). Competence and accumulated skills and knowledge affect the path-dependent capabilities within an LIC to create new routes to competitive advantage.

4. THE EVOLUTION OF THE MUSICAL INSTRUMENT CLUSTER OF ANCONA

Knowledge creation through the acquisition of information and the adaptation of mobilized local capabilities to new conditions requires in some cases high-tech investments by incumbent firms. Although the initial high-tech investments in the 'accordion' cluster of Ancona were mostly carried out autonomously by single incumbent firms, the entire LIC benefited from the creation or transfer of knowledge deriving from those investments. These 'leading-edge investments' (Malmberg and Maskell 2001) in most cases led to a severe crisis for the investing firm. However, the knowledge created by these investment projects allowed the LIC to develop. This was fostered by the mobility of employees, who moved to other firms or founded new firms. The need to satisfy the local demand for technologically advanced components by the local musical industry led some local sub-suppliers to adopt high-tech production processes and products. These firms absorbed existing knowledge coming from other industries. Most of the firms to follow this strategy were incumbent firms with strong links to external networks. This allowed them to serve different markets with their products and to differentiate their customer portfolio.

The process of change in the production structure was achieved by integrating new products into the production portfolio. This led to a quite different relational structure within the LIC and to the creation of new production niches. The LIC is now formed by several different local production networks still connected with, but different from, the original accordion network. Once the information about market opportunities and threats was available, firms sought new production niches. The set of possible change for individual firms depended on the internal learning process and the absorptive capacity of the LIC. Start-up firms specializing in the production of technological devices developed. This strengthens Almeida and Kogut's (1997, p. 23) observation that local production and knowledge networks 'decrease the uncertainty and costs associated with start-up activities'. These start-ups created new local knowledge networks that specialized in electronics. The 'natural' interrelation of the electronic musical instrument industry with other industries facilitated the creation of a certain degree of variety, which led to the discovery of new production niches and to a structural change within the LIC.

Overall, the transition towards electronics within the LIC dramatically reduced the number of firms in accordion production, but also the number of firms in electronic instrument production. Table 10.1 shows in terms of firms and employment how the musical instrument industry lost importance within the LIC. To give an indication of the importance of the musical instrument industry, note that Castelfidardo alone still hosts about 100 firms and 1200 employees in the industry.

The next two sections will explore in more detail the different roles played by some of the firms. Different firms take up different roles in absorbing, processing and distributing knowledge among local actors, confirming the observation made by Rabellotti and Schmitz (1999) which emphasized the different roles of firms in terms of size, performance and the 'level of local embeddedness' in the footwear clusters in Italy, Brazil and Mexico. The information was obtained through about 20 semi-structured interviews with entrepreneurs, managers and other local actors. The musical instruments cluster emerged in Castelfidardo (Ancona), in the Marche region (Italy), but then extended its boundaries to Osimo, Camerano, Loreto, Recanati, Potenza Picena and Montecassiano, which are all municipalities of the same region.

Table 10.1 The development of the musical instrument industry in the region of Ancona (in % of total industry)

	1961		1971		1981		1991	
	Firms	Employees	Firms	Employees	Firms	Employees	Firms	Employees
Ancona	1	1	2	0	4	4	0	0
Camerano	9	60	5	61	4	27	2	11
Castelfidardo	54	90	40	67	31	44	24	19
Loreto	10	76	7	12	25	39	2	1
Osimo	6	22	5	10	9	16	65	7
Sirolo	0	0	19	39	21	36	1	1

Source: ISTAT.

4.1 The Transition to Electric and Electronic Instruments

> Castelfidardo, Italy – Nothing has been the same in this little town since the Beatles started turning young men's thoughts to electric guitars. (*The New York Times*, 21 May 1988).

The interest in electric and electronic music grew during the 1960s and 1970s and the demand for electric and electronic instruments increased. Production attempted to keep up with the trend. The introduction of electric valves to instruments in the 1960s and electronic transistors in the 1970s does not in itself represent a disruptive phase for the LIC. On the contrary, it marked the beginning of a successful period. The same is not true for the single innovating firms, some of which did not survive the transition. Many firms needed knowledge to integrate electronics into their products. They hired engineers and technicians from northern Italy. The strategy adopted by one incumbent firm (Farfisa) in order to organize the transition to electronic musical instruments was to temporarily employ electronics experts who worked in close relationship with local skilled labor. This strategy aimed at shortening the time needed to absorb technical knowledge and make it operative. During this period, a process which Belussi and Pilotti (2000) call 'contextualization of external (codified) knowledge' took place. Technical codified know-how in electronics was directly integrated into the local musical instrument production. In the meantime, accordions began going out of fashion and the producers could not compete on price with competitors from East Germany, Czechoslovakia and China.

The production process was turning from artisanry into one dominated by technology. Almost all firms followed this change. At the beginning of the transition some incumbent firms invested directly in electronics. Electronic supports simulating other instruments were added to the traditional accordion, which made it possible for a sole musician to play like an orchestra. The transfer of technological knowledge from different external technological knowledge bases to the musical industry in the LIC was deepened by new firms specializing in electric and electronic components and producing musical instruments on the basis of the new technologies. A wave of new start-ups took place. Between 1961 and 1971, 25 new firms producing musical instruments were set up in the area of Ancona and 32 in the area of Macerata. In the following decade between 1971 and 1981, 14 new firms entered the area of Ancona and 121 the area around Macerata. In 1981 the musical instrument firms in the district area employed 2173 persons.

How could the small and primarily locally embedded firms access external codified knowledge? The interviews with the managers of larger local firms (in the music or electronics production sectors or in the service sector) revealed that these firms were not only embedded in local networks but were also part of non-local networks of production and exchange of knowledge. They attributed more importance to the non-local network as regards the identification of technological and market opportunities. The non-local networks were mostly related to the size and organization of the firm. Non-local branches were especially important in this respect. These also gave access to informal participation in other communities of practice and networks (Amin 2000). These non-local connections were established during the first expansion phase of the LIC, when leading firms located new branches outside the area, and more recently by service providers which are establishing long-term relationships with other service providers and in some cases are providing service in different LICs. The interviews showed clearly that start-ups had to rely almost exclusively on the local networks. Those small firms had fewer means to build up external contacts. However, with the passing of time they profited from the activities of the large firms, often in terms of vertical relationships.

Most larger firms invested in electronics during the 1980s. However, firm-specific events prevented investment into electronics. For example, Elka had just invested in a new building at the time and could not afford to hire engineers or invest in electronics. In 1988 Elka was absorbed by General Music (GEM), a firm external to the district that had invested heavily in electronics and was looking for expansion and skilled labor. Today General Music specializes in digital keyboards and has around 400 employees, four plants and affiliated companies in Britain and in the Netherlands. The main production plant is located in Recanati, within the district area. However, the administrative center is in San Giovanni in Marignano (Rimini, Emilia Romagna region). Apart from personnel working in design and R&D, the skilled labor employed at GEM's plant in Recanati is local. Most employees had been working for Elka. The reason for locating the production plant in the major agglomeration of firms producing musical instruments is the concentration of skills, capabilities and knowledge in the area. This demonstrates Marshall's dictum that 'in all but the earliest stages of economic development a localized industry gains a great advantage from the fact that it offers a constant market for skill' (Marshall 1920). Once a critical mass of specialized firms has been reached, it becomes convenient for other firms producing the same or similar products to locate their production facilities in the same location, as they can benefit from external economies. In the case of GEM, the engineers working in San Giovanni in Marignano do the R&D and develop new ideas for products. However, the practical implementation of the ideas into new products takes place at the plant in Recanati within the LIC.

Although high-tech investments in the 'accordion' cluster of Ancona (Italy) have been mostly carried out autonomously by single larger firms, the entire LIC benefited from the creation of knowledge derived from those investments. In this respect it is interesting to compare the fate of two different firms still both producing accordions, Borsini and Cemex. The Borsini enterprise (now in the third generation of family ownership) has continually improved its accordions by preserving the traditional product and refusing radical changes. Even when changing demand forced the introduction of electronic parts to the traditional accordion, Borsini relied on external firms to insert these parts. The other firm, Emex/Excelsior (now Cemex), began in the USA (New York) when a local emigrant recognized the possibility of transforming the artisan tradition of his hometown (Castelfidardo) into a successful enterprise. Emex attempted the transition to electric and electronic production by employing engineers and technicians, but it did not succeed. Ex-employees took over the firm and changed the name to Cemex. Cemex is now successful in producing traditional accordions for the international market.

4.2 Start-ups and Electronics

A second phase of the transition in many cases saw the strengthening of a tendency toward externalization. The pioneering exploration of new markets for known technologies by leading firms allowed incumbent firms to differentiate their products, and made the start-up of electronics firms possible. During the transition of the LIC towards electronics a number of engineers employed previously by firms producing musical instruments set up small firms specializing in electronic components for musical instruments. Now the area counts several firms specialized in electronics and electromagnetics, producing components for music instruments as well as for other sectors including automobiles, security systems and telecommunications.

The story of firms which produce electronic home appliances is interesting. Some firms diversified their production and specialized in electronics. These firms enlarged their production range on the basis of skills and knowledge accumulated in moving from musical instruments to electronic musical instruments. By further developing the techniques used to make amplifiers, these firms were able to enter the video, telephone and video-control systems industries. Some of these firms even discontinued musical instrument production altogether. This is the case with Farfisa, a firm founded in 1946 by Paolo and Settimio Soprani. Farfisa soon became one of the largest firms in the musical instrument industry in Italy with over 1000 employees by switching promptly into electric and electronic musical instruments. In 1967 Farfisa decided to concentrate its electronic knowledge in Farfisa Intercom. Following this step Farfisa developed its R&D capabilities by employing about one hundred researchers in the 1970s. The aim was to figure out alternative or complementary products based on the already existing technological expertise. At this time some workers with traditional skills exited the firm and founded their own musical instrument enterprises. But Farfisa also experienced spin-offs from the other side. In 1972 the head of the R&D division founded a start-up in Ancona. This venture, Aethra, is now one of the world leaders in video conferencing, ISDN systems, equipment for application in telecommunication and data transmission. The firm's most important source of information and learning is its local R&D division and their local and non-local networks of partners and resellers. However, Farfisa's products portfolio also evolved towards the fields of video intercoms, telephone and video-control systems, and in 1993, Farfisa was split up in two independent businesses by a management buy-out. The communication and security business was separated from that of musical instruments. The former took over the brand name, most of the know-how, personnel and plants. One of the most famous electric and electronic music instrument companies no longer produces musical instruments.

This case shows how firms can reinvent themselves along a trajectory of technological knowledge so that they end up producing products which have nothing in common with the products produced at the beginning of their life.

Aside from this example of a large firm managing the transformation of its product portfolio, there are several small subcontractor firms specialized in electronics producing transistors and other electronic components which are suitable for musical instrument, home electronic appliances and even personal computer production. Some of those small start-up firms specialized in the production of transistors and computer components. The technological knowledge base for the start-ups was the technology of producing components for electronic keyboards. Today the LIC includes a small new network of integrated-circuit-producing firms. Some of these firms are among the largest in Italy in respect of export volume in this industry. However, most of these firms are very small, and they are much more dependent on the local environment than are the larger firms. They are locally rooted in respect of both finance and knowledge. Indeed, most start-up firms are financed by the local credit system. The small electronic firms rely mostly on common local knowledge sources. Due their small size, these firms often develop joint projects. This creates a network of exchange of information and knowledge. The high mobility of personnel in the area, particularly among start-up firms, makes knowledge transfer very easy.

Interviews revealed that the University of Ancona has a central role. Its alumni are an important source of knowledge for these firms. The extent of product differentiation and served markets attained by these firms is relatively high. In addition to supplying the musical instrument industry, the electronics firms supply the automobile industry with printed circuits, the telecommunication industry with ISDN devices, video communication and audio-conference systems, cameras, minicameras and monitors.

4.3 Changes in the Organizational Form of Firms

Firms have changed not only their production focus but also their organizational forms over time. During periods in which suppliers with market power imposed high prices, some firms chose to integrate vertically by internalizing those activities. Larger firms in particular acquired suppliers and integrated vertically. Some of the firms reallocated these activities to the outside when convenient. Especially in recent times the presence of only a few suppliers of crucial components within the LIC persuaded many firms to adopt a strategy of vertical integration. However, in most of the cases organizational integration took place because of knowledge complementarities which required better coordination of the different capabilities in the process of systemic innovation. The transaction costs involved were primarily 'cost(s) of persuading, negotiating with, coordinating among and teaching outside suppliers in the face of economic change or innovation' (Langlois and Robertson 1996; see also Malerba and Orsenigo 2000). This shows that what has been regarded as the most resistant and successful typology of an LIC, the 'Marshallian industrial district', evolved towards a different form of organization, as not only the organizational form of firms changed but also the organizational form of the cluster. Organizational form changes over time and adjusts to the technological path and to other conditions (see, for example, Lazerson and Lorenzoni 1999, Franchi and Rieser 1991, Piore and Sabel 1984).

Firms which integrated other products, after having started electronic instrument production, explored the possible options requiring similar capabilities as those involved in the electronic instrument branch (for a general discussion on this mechanism see Loasby 1996). It was the degree of 'modularity' in competencies, skills and technology that guided the process of change within the LIC. For the modularization of the competencies the heterogeneity of firms was instrumental, as these organized the translation of external technological knowledge into localized competence. The case study shows that different organizational forms can be identified in the LIC and that some of them have changed over time. Moreover, it shows that while the underlying technologies and core competencies of industrial regions are of crucial importance, it is the heterogeneity of behavior, strategy and organizational forms that guides changes in localized core competencies.

5. DISCUSSION

During the 1960s and 1970s the firms within the LIC of Ancona organized a process which transformed the traditional industrial district of musical instruments into a multi-layered localized industrial cluster containing traditional accordion producers, electronic musical instrument producers and electronics firms. Today the LIC is made up of several (more or less) interrelated networks. The musical instruments network produces approximately 80 percent of Italian exports in musical instruments. And even if mass music 'demand' has turned to electronic instruments, accordion production still enjoys a niche demand which can be satisfied by firms in the Marches cluster still producing accordions (for example Cemex). In these firms the abilities, skills, and tacit competencies of the original district have been kept alive. Beside those a specialized electronics industry emerged within the LIC. First the firms in the electronics sector supplied almost exclusively the local musical instrument producers; later this industry got a life of its own.

This process of change is best conceptualized as a process of self-organization (see Foster 1997, Witt 1997, 1998) carried out by local forces. The process was triggered by external events and changes in strategies of larger firms, which had the ability to access external technological knowledge and information channels. These firms integrated the new technology (electronics) into their products. Through employee turnover and spin-offs this new technological knowledge was embedded into the local economy and led to the formation of new networks, less and less connected to the old industrial district. The transition was initiated by established larger firms, but it was the small local start-ups that realized the reorganization of the LIC. The case study showed that production systems 'rarely develop successfully by simply producing and selling what they happened to discover is demanded by experience' (Foster 1997). On the contrary, firms must perceive new market opportunities and explore new production niches. However, this process of reconfiguration must be considered as the result of the peculiar historical and technological path of the LIC. The key issue cannot be reduced to the adoption of the right technology. However, some elements can be isolated.

One of the central elements for the transformation of the LIC was the presence of heterogeneity in behavior, strategy and organizational form.

Larger firms were instrumental in bringing into the local economy knowledge about new products and new technologies. These pioneering incumbent firms thereby created niches for start-up firms supplying components. Even though those firms exploring new market opportunities were less locally reliant and were able to build durable relationships outside the LIC, they were instrumental in bringing in novelty and connecting the local dimension with the outside world. However, to have an impact a locally based collective learning process had to come into existence, whose absence may lead to the decline of an LIC. The large number of spin-offs clearly demonstrates that the high mobility of employees within the LIC's area implemented the know-how about electronics into an collective local learning process. It was the small start-up firms which embedded the new technological knowledge into localized competence. This shows that an environment which 'promotes collective learning and flexible adjustment among specialist producers of a complex of related technologies' (Porter 1998) can cushion in an efficient way the risks of entrepreneurial activity. The increasing local demand for electronic components by musical instrument producers created an environment which reduced the risks and costs of start-ups, and fostered the use of new technologies.

NOTE

* I wish to thank all the local interviewees, in particular Marco Galeazzi (Cemex) and Vincenzo Borsini for devoting precious time to my curiosity, Mattias Weber and the participants in the EMAEE 2001 in Vienna, as well as an unknown referee, for their helpful comments, Peter Maskell and the members of the Department of Industrial Economics and Strategy at the Copenhagen Business School for interesting discussions. The usual disclaimer applies.

REFERENCES

Almeida, P. and Kogut, B. (1997) 'The exploration of technological diversity and the geographic localization of innovation', *Small Business Economics*, **9**, 21–31.

Amin A. (2000), 'Organisational learning through communities of practice', paper presented at the 8th Millennium Conference of the Schumpeter Society, University of Manchester, 28 June–1 July.

Antonelli, C. (2000), 'Collective knowledge communication and innovation: the evidence of technological districts', *Regional Studies*, **34**, 535–47.

Belussi, F. and Pilotti, L. (2000), 'Knowledge creation and collective learning in the Italian local production systems', University of Padua, Working Papers.

Boschma, R.A. and Lambooy, J.G. (1999), 'Evolutionary economics and economic geography', *Evolutionary Economics*, **9**, 411–29.

Brenner, T. (2000), 'The Evolution of Localised Industrial Clusters: Identifying the

Process of Self-Organisation', *Papers on Economics and Evolution* No. 11, Jena: Max Planck Institute for Research into Economic Systems.
Breschi, S. (2000), 'The Geography of Innovation: A Cross-sector Analysis', *Regional Studies*, **34**, 213–29.
Camagni, R. (1991), *Innovation networks: Spatial perspective*, London: Belhaven Press.
Cohen, W.M. and Levinthal, D. (1989), 'Innovation and learning: the two faces of R&D', *Economic Journal*, **99**, 569–96.
Cohen, W.M. and Levinthal, D. (1990), 'Absorptive capacity: A new perspective on learning and innovation', *Administrative Science Quarterly*, **35**, 128–52.
Dosi, G. (1984), *Technical change and economic transformation*, London: Macmillan.
Dosi, G., Freeman, C., Nelson, R., Silverberg, G. and Soete, L. (eds) (1988), *Technical Change and Economic Theory*, London: Pinter Publishers.
Foss, N. (1997), 'Evolutionary Economics and the Theory of the Firm: Assessments and Proposals for Research', in J. Reijnders (ed.), *Economic and Evolution*, Cheltenham, UK and Lyme, USA: Edward Elgar.
Foster, J. (1997), 'The analytical foundation of evolutionary economics: From biological analogy to economic self-organization', *Structural Change and Economic Dynamics*, **8**, 427–51.
Foster, J. and Wild, P. (1999), 'Detecting Self-Organizational Change in Economic Processes Exhibiting Logistic Growth', *Journal of Evolutionary Economics*, **9**, 109–33.
Franchi, M. and Rieser, V. (1991), 'Le Categorie Sociologiche nell'analisi del Distretto Industriale! Tra Comunita le Razionalizzazione', *Stato e Mercato*, **33**, 451–76.
Fujita, M. (1996), 'On the Self-organization and Evolution of Economic Geography', *Japanese Economic Review*, **47** (1), 34–61.
Guerrieri P., Iammerino, S. and Pietrobelli, C. (2001), *The Global Challenge to Industrial Districts*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Johnson, B., Lorenz, E. and B.–A. Lundvall (2001), 'Why all this fuss about codified and tacit knowledge?', *Industrial and Corporate Change*, **11** (2), 245–62.
Langlois, R.N. and Robertson, P.L. (1996), 'Explaining vertical integration: Lessons from the American automobile industry', in M. Casson (ed.), *The Theory of the Firm*, Cheltenham, UK and Lyme, USA: Edward Elgar.
Lazerson, M.H. and Lorenzoni, G. (1999), 'The firms that feed industrial districts: A return to the Italian source', *Industrial and Corporate Change*, **8**, 235–66.
Lissoni, F. and Pagani, M. (2002), 'How Many Networks in a Local Cluster? Textile Machine Production and Innovation in Brescia', in T. Brenner and D. Fornahl (eds), *Cooperation, Networks and Institutions in Regional Innovations Systems*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Loasby, B.J. (1996), 'The organization of industry and the growth of knowledge', Lectiones Jenense 7, Max Planck Institut zur Erforschung von Wirtschaftssystemen, Jena, p. 23.
Lombardi, M. (2001), 'Cognitive equilibria, efficiency, and discontinuities in the evolution of Local Production Systems', paper presented at the Congress Sistemi Governance & Conoscenza nelle Reti di Impresa, University of Padua.
Malerba, F. and Orsenigo, L. (2000), 'Knowledge, Innovative Activities and Industrial Evolution', *Industrial and Corporate Change*, **9**, 289–305.
Malmberg, A. and Maskell, P. (2001), 'The elusive concept of localization economies: Towards a knowledge-based theory of spatial clustering', paper presented at the AAG Annual Conference, New York, February.

Marshall, A. (1920), *The Principles of Economics*, London: Macmillan.
Metcalfe, J.S. (1998), *Evolutionary Economics and Creative Destruction*, London: Routledge.
Nelson, R. and Winter, S. (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
Orillard, M. (1997), 'Cognitive Networks and Self-organization in a Complex Socio-economic Environment', in A. Amin and J. Hausner (eds), *Beyond Market and Hierarchy: Interactive Governance and Social Complexity*, Cheltenham, UK and Lyme, USA: Edward Elgar.
Piore, M. and Sabel, C. (1984), *The second industrial divide*, New York: Basic Books.
Porter, M.E. (1998), 'Clusters and competition', in M.E. Porter, *On competition*, Cambridge, MA: Harvard Business Review Books.
Rabellotti, R. and Schmitz, H. (1999), 'The Internal Heterogeneity of Industrial Districts in Italy, Brazil and Mexico', *Regional Studies*, **33**, 97–108.
Saviotti, P. (2001), 'Considerations about a production system with qualitative change', in J. Foster and J.S. Metcalfe, *Frontiers of Evolutionary Economics: Competition, Self-Organization and Innovation Policy*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Tappi, D. (2002), 'On the unit of analysis in the study of networks', in T. Brenner and D. Fornhal (eds), *Cooperation, Networks and Institutions in Regional Innovation Systems*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Witt, U. (1997), 'Self-organization and Economics – What is New?', *Structural Change and Economic Dynamics*, **8**, 489–507.
Witt, U. (1998), 'Coordination of Individual Economic Activities as an Evolving Process of Self-organization', in P.J. Boettke and D.L. Prychitko (eds), *Elgar Reference Collection, International Library of Critical Writings in Economics*, Cheltenham, UK and Northampton, USA: Edward Elgar.

11. Evolutionary thinking in environmental economics: retrospect and prospect

Jeroen C.J.M. van den Bergh*

1. INTRODUCTION

This chapter explores the various links between environmental and evolutionary economics. Environmental economics combines elements of natural and social sciences, and touches upon basic research questions. At the same time, it is policy-oriented. The combination of basic research and policy relevance is exciting. Within environmental economics it is almost 'natural' to think about similarities and possible links between biological and social–economic evolution. The result is 'co-evolution' that is very 'real' and concrete. One would expect that social sciences in particular can benefit from such a transfer, given that evolutionary biology has a long tradition. Nevertheless, one should certainly not exclude the possibility that learning can occur in the other direction. Biology in particular can learn from evolutionary theory and modelling in social sciences when it comes to analysing the evolution of groups and social organization.[1]

The relevance of evolutionary thinking to understanding environmental problems and formulating public policy responses can be illustrated in several ways. Environmental problems often mean a loss of diversity of options, which is best illustrated by the loss of biodiversity. Many environmental problems are difficult to resolve because undesirable, second-best technologies are locked in as a result of historical accidents and increasing returns to scale. The most significant example is the complete dependence of modern economies on fossil fuels. Related to this is the dominance in both passenger and freight transport of cars with fossil fuel combustion engines. An entirely different example of the relevance of evolution for environmental economics is that many resource-based sectors, such as agriculture, fisheries and forestry, affect not only the size of renewable resources and ecosystems, but also their internal (genetic) diversity. In the long run this can undermine the profitability of these activities.

Although the relevance of evolutionary thinking in environmental economics is clear, the diffusion of it has been very slow. Nevertheless, in the last few decades a small number of authors have tried to introduce evolutionary aspects into environmental economic analysis. The major contributions can be summarized as follows.

Georgescu-Roegen (1971) discussed exosomatic instruments as an almost natural extension of endosomatic capabilities of humans – that is, those which are part of the human body. Of all authors, Boulding (1978, 1981) is the one who has most clearly and consistently emphasized the analogy between ecology, evolutionary biology and economics, focusing on concepts such as homeostasis, population and the distinction between genotype and phenotype. Norgaard (1984) was the first to discuss seriously the application of the concept of co-evolution to the interaction between economic and natural systems. Ayres (1994) is a very original synthesis in book form of evolutionary ideas about virtually any aspect of reality, including economics and environment. Gowdy (1994, 1999) has stressed that evolutionary biology is not only about micro-evolution, but also incorporates elements of macro-evolution. This may suggest that economic evolution operates at multiple levels as well (see also van den Bergh and Gowdy 2002). Faber and Proops (1990) have combined elements of the neo-Austrian approach – with its emphasis on the temporal and roundabout features of economic processes – and evolutionary thinking, very much based on the distinction between genotype and phenotype. Munro (1997) has offered a first analysis that combines net present value maximization of agricultural crop cultivation with the effects of selective pressure of pesticides on the genetic composition of a pest population. Recently, Jackson (1999, 2000 and 2002) has argued that if ecological economics seeks 'consilience' with biology, then it must incorporate evolutionary theories and modelling. Finally, Sethi and Somanathan (1996) have considered the endogenous character of norms in common-pool resource situations with evolutionary game theory. Some of these studies are discussed in more detail in subsequent sections.

Apart from the relevance of evolutionary thinking for the study of environmental problems and their policy responses, another important link between economic evolution and the environment needs to be mentioned. This is especially relevant for the core of evolutionary economics. So far, economic evolutionary theories have completely neglected natural environmental and resource dimensions. This is understandable. The neo-Schumpeterian school has very much focused on technological innovation over relatively short time horizons, at least from a biological evolutionary angle. And evolutionary game theory has focused on extremely simple mathematical models that are analytically solvable. Interestingly, these models mainly focus on selection dynamics; that is, they completely ignore innovation and, worse, the interaction between selection and innovation, which makes evolutionary processes so unique and unpredictable at a detailed level. From this perspective, the name suggested by some researchers in this area, 'equilibrium selection theory', is more appropriate than 'evolutionary economics'.

When studying long-term economic history and future within an evolutionary framework, neglecting environmental and resource dynamics may result in a very biased model of reality. The reason is that important phases of economic history emerged under the strong influence of environmental and resource factors. This in any case holds for such major transitions as the development of agriculture and human settlements, and the Industrial Revolution. Energy scarcity and climate change may have been important triggering factors. More will be said about this later on.

This chapter is organized as follows. Section 2 presents a short overview of environmental economics, so as to accustom the reader to its core concepts and approaches. Attention will be devoted to 'ecological economics', which is argued to be closer in spirit to evolutionary economics. Section 3 addresses the relationships and possible conflicts between economic growth and environmental quality from an evolutionary angle. This involves considering evolutionary growth modelling and identifying positions in the 'growth debate'. Section 4 takes an even longer temporal perspective, by discussing the role of environmental resources in major structural changes in human-economic systems. Section 5 presents a set of weakly related insights derived from evolutionary analysis at the ecosystem and single resource level. Section 6 identifies elements of an evolutionary theory of environmental policy. Section 7 concludes.

2. A CONCISE OVERVIEW OF ENVIRONMENTAL ECONOMICS

Environmental economics is a branch of economics that is concerned with the economic analysis of the origin and nature of environmental problems, as well as their solutions. This includes issues relating to markets as well as to public policy. Environmental economics is often defined as including resource economics. The combination makes sense, since many resource issues are intricately linked to environmental issues. This is perhaps most noticeable in the case of fossil energy resources, the use of which contributes to the enhanced greenhouse effect. The notion of sustainable development, which has become one of the pillars of modern environmental economics, underpins the linkage of resource and environmental problems.

Environmental economics developed out of applied welfare economics influenced by agricultural and resource economics during the 1960s. The early development of environmental economics was dominated by three themes. First, cost–benefit analysis was applied to investment projects with environmental impact, notably related to water (rivers). Closely related to this, monetary valuation techniques were developed and applied to value environmental changes and damage. Second, environmental policy theory was developed, aimed at the evaluation, comparison and design of environmental policy instruments. Third, economic growth and resource scarcity were examined in theoretical and empirical studies. This was linked to resource economics, which covered a number of issues including the testing of resource scarcity – with a range of indicators reflecting physical conditions, costs or prices – optimal resource extraction, imperfect resource markets, extraction of non-renewable resources such as fossil fuels, metal ores and minerals, and use and management of renewable resources such as water, forestry, fisheries, wind and solar energy.

Core themes in modern environmental economics include environmental policy theory, international environmental problems, growth and sustainable development (macroeconomics), and monetary valuation (see van den Bergh 1999). Recently an area known as 'environmental (business) management' has also received much attention. It has strong links with environmental economics, because it adopts a business perspective to understand the firm's responses to environmental problems and policies. This field of research covers typical business administration topics, such as environmental care systems, environmental strategies, internal organization, environmental accountancy, environmental reporting, environmental cost accounting and green marketing.

The economic theory of environmental policy starts from the concept of externalities, which can be defined as the influence of one economic agent's decision on the utility or production of another agent that occurs outside the market and remains uncompensated (Baumol and Oates 1988). The presence of externalities means that individuals do not have complete control over the set of factors that determines their production or utility levels. Environmental economics is particularly interested in negative environmental externalities, such as the negative physical effects of environmental pollution, resource use, or other types of environmental disturbance – such as fragmentation due to road infrastructure in nature areas – by one agent on another. Externalities have been analytically examined and elaborated with the help of partial and general equilibrium theories, consistent with neoclassical assumptions regarding individual behaviour and operation of markets.

Instruments of environmental policy are traditionally evaluated in economics on the basis of their efficiency features. Effectiveness and distribution effects (equity, fairness) function as secondary evaluation criteria. The most common (archetypal) comparison is between uniform standards and taxes in pollution control. Taxes are attractive as they provide better incentives than standards to change individuals' behaviour, and thus realize more efficient outcomes: either social welfare is higher or costs of realizing fixed targets are lower. They do this by equalling marginal costs of pollution abatement, assuming that individual polluters are minimizing costs. The best-known price instrument of environmental policy theory is the optimal or Pigouvian tax, defined as equal to the marginal external costs in the optimal equilibrium. This optimal tax can be adapted for imperfect markets, dynamic contexts (technological innovation) and transaction costs. Standards are especially attractive from the perspective of effectiveness (or uncertainty). A combined instrument is a system of tradable permits. This has two features: a ceiling is set on all pollutive emissions of a particular type by granting a finite amount of emission permits; and permits are tradable. The first feature makes sure that the total (national, regional) emission level is restricted; the second feature ensures flexibility and efficiency at the level of individual agents, with the ideal result that costs of pollution abatement at the margin are equal among all firms and equal to the equilibrium permit price. From an evolutionary perspective price instruments such as taxes and tradable permits are interesting, as they reflect a public regulatory response to the presence of heterogeneous firms.

Since the late 1980s, in the wake of the popularity of the notion of sustainable development, international dimensions of problems and policy have received much attention in environmental economics. A first question here is how environmental policy in open economies can be best designed. This depends very much on the type of problem: local (solid waste, urban pollution), transboundary ('acid rain', river pollution) or even global (greenhouse gas emissions and climate change). Another consideration is whether pollutive goods are traded, in relation to whether pollution relates to the production or the post-consumption stage. Many theories have been used to clarify the link between foreign trade, environment and policy based on: partial equilibrium, imperfect competition, strategic trade policy by countries, general equilibrium, statistical–econometric. Special attention has been paid to the role of developing countries and the very large differences in technology, income level, and institutions between rich and poor countries.

Environmental macroeconomics is a term that refers to the macro-scale of environmental issues, involving especially the impact of economic growth and the goal of sustainable development (see van den Bergh 1999, part VI). The environment has received much attention in macroeconomic modelling because economists have since long been asked to provide information about the national economic consequences of environmental policy plans. The theory of growth has been used to clarify relationships between growth, resources and pollution. At the moment, research in this area is dominated by endogenous growth theory, even if alternative approaches have been proposed (van den Bergh and Hofkes 1998).

Economic valuation is perhaps the area where environmental economists have truly generated their own theory (Freeman 1993). The idea of valuation theory is that monetary valuation can be performed using monetary compensation or equivalent measures, which involve comparing two states, one before and one after a certain environmental or policy change. Monetary valuation is done using several concrete methods. Travel cost models try to assess the value of nature areas or parks by linking recreation demand to generalized individual travel costs, which include the value of time spent in travelling. Hedonic price models work similarly, providing a relation between property (land, houses) on the one hand and environmental indicators on the other hand. Both these methods are said to 'reveal preferences'; that is, they use information about actual choices made by individuals in existing markets. Another method, contingent valuation, has perhaps received the most attention. It can be relatively easily applied. It is based on stated preferences, that is, asking individuals directly for their value associated with a hypothetical change scenario. The main advantages of this method are that it can address a wide range of (hypothetical) environmental changes, including those that have not occurred, and that it can assess non-use values. The main disadvantage is that it suffers from various biases due to the hypothetical nature of the questions. In addition, the method is very sensitive to design characteristics. Economists have debated correct procedures for this method, especially following the valuation studies done in response to the *Exxon-Valdez* oil spill in Alaska in 1989 (see Carson et al. 1992; Hausman 1993; Arrow et al. 1993). Other methods for valuing environmental change can be based on production function techniques, notably in resource sectors such as agriculture and fisheries, and on adopting theoretically less attractive measures such as defensive expenditures, relocation costs, restoration costs and opportunity costs. Since the 1990s interest has increased in two related methods, namely meta-analysis (van den Bergh et al. 1997) and value transfer (Brouwer 2000). Meta-analysis can be defined as the formal, statistical synthesis of results and findings of scientific studies. It can be used to summarize relationships and indicators over a collection of similar studies, to compare different methods applied to similar questions, and to trace factors responsible for differing results across similar studies. It can serve as the methodological basis for value transfer, also known as benefits transfer, which is aimed at transposing monetary values from one site to another. This is regarded as cost-effective relative to performing a new 'primary valuation study'. Transfer is controversial, however, because it should satisfy a number of criteria, including: adequate data, sound economic methods, similar populations of study and transfer sites, and similar site characteristics. Economic valuation can be criticized from many angles (see Gowdy 1997, Blamey and Common 1999). The most fundamental criticism focuses on the standard assumptions of economic valuation, arguing that notions of non-utilitarian altruism, ethical attitudes and 'citizen responses' mean that individual behaviour cannot be reduced to maximizing a given and stable utility function (van den Bergh et al. 2000).

The previous concise overview illustrates that traditional environmental and resource economics is very much dominated by neoclassical microeconomics. This is most clearly exemplified by theories of monetary valuation and environmental policy. Since the end of the 1980s, along with the increased attention to sustainable development, 'environmental macroeconomics' and multidisciplinary analysis have received more attention from economists (van den Bergh 1999). Much of this work goes under the heading of 'ecological economics'. This was founded as a new field of research at the end of the 1980s. It has been quite successful in an institutional sense, as shown by the following indicators: a much-cited journal with the same name, many book publications, regular conferences and workshops, and last but not least, an international society with regional chapters.

Ecological economics (EE) is a more academic and social-science-oriented version of environmental science. It integrates elements of economics, ecology, geography, political science, thermodynamics, ethics, and various other natural and social sciences. The core of ecological economics can be associated with the goal of sustainable development, interpreted as both intra- and intergenerational equity. It supports the view that the economy is a subsystem of a larger local and global ecosystem that sets limits to the physical growth of the economy. It is characterized by methods that use physical – material, energy, chemical, biological – indicators and comprehensive systems analysis. Ecological economics provides a forum for multidisciplinary environmental research as well as an alternative view and approach to traditional environmental (and resource) economics. Nevertheless, ecological economics has perhaps been most successful in promoting multidisciplinary research in which natural scientists and social scientists join forces. Gowdy and Ferrer-i-Carbonell (1999) refer in this context to the ideas by the famous biologist E.O. Wilson (1998) on 'consilience' between biology and economics as an extended form of consistency. Indeed, the economistss K.E. Boulding, H.E. Daly and

Table 11.1 Differences in emphasis between evolutionary, ecological and traditional environmental and resource economics

Evolutionary environmental economics	Ecological economics	Traditional environmental and resource economics
Evolutionary potential	Optimal scale	Optimal allocation
Variety	Equity	Efficiency
Evolutionary stable strategies	Sustainable development	Sustainable growth
Adaptive limits	Limits to growth	Growth of limits
Path-dependence	Irreversibility	Quantifiable uncertainty
Very long run	Long run	Short/medium run
Adaptive systems	Causal processes	Abstract statics/ dynamics
Population/distribution indicators	Physical and biological indicators	Monetary indicators
Bounded rationality	Myopic behaviour	Rational behaviour
Fitness counts	Environmental ethics	Utilitarianism
Open system (energetically)	Open system	Closed system

N. Georgescu-Roegen, and the ecologists C.S. Holling and H.T. Odum can be considered as intellectual founders and antecedents of ecological economics. Interestingly, some of these have subscribed to evolutionary approaches to economics (see Section 1). This is consistent with evolutionary and ecological economics having a number of similar starting points. Both are built upon (population) ecology, limits and adaptive systems, diversity of agents and bounded rationality, and a focus on causality and history (see Table 11.1 for a more systematic comparison). See van den Bergh (2001) for a short introduction to ecological economics as well as a comparison with traditional environmental economics.

3. EVOLUTIONARY ANALYSIS OF GROWTH, ENVIRONMENTAL LIMITS AND PROGRESS

In this section we will consider the relationship between economic growth, environment and resources, which has always been an important motivation for environmental economic analysis from the perspective of evolutionary thinking. The old 'growth debate' (van den Bergh and de Mooij 1996) can be characterized by three core questions: Is growth desirable? Is it feasible? And can it be controlled or steered? By simply checking all combinations of yes/no answers to these questions one can identify eight posi-

Table 11.2 Possible perspectives in the growth debate

Perspective	Core questions			Interpretation/Implication
	Growth desirable?	Growth feasible?	Growth controllable?	
1	yes	yes	yes	Continuous growth is possible: *extreme optimist*
2	yes	yes	no	Objective of continuous growth is perhaps unreachable: *optimist*
3	yes	no	yes	Objective of continuous growth is unreachable, but disaster can be avoided: *pessimist*
4	yes	no	no	Objective is unreachable, and disaster is possible: *extreme pessimist*
5	no	yes	yes	Undesired growth, but it can be controlled: *immaterialist*
6	no	yes	no	Possibly undesired and uncontrolled growth: *opportunist*
7	no	no	yes	Growth is not necessary, not possible, and can be controlled: *immaterialist and pessimist*
8	no	no	no	Undesired growth can result in a disaster: *immaterialist and extreme pessimist*

tions in the 'growth debate'. These are listed in Table 11.2, along with short interpretations and implications.

Which answers to these questions result from an evolutionary perspective? The standard economic view on sustainable growth, focusing on gradual growth in equilibrium, is not shared by evolutionary economics. Selection and mutation processes will continue to change the structure of the economy, characterized by irreversible changes in the diversity of competing alternatives which include processes, products, firms, individuals, groups and regions. As argued in the previous section, as opposed to economic growth theory, evolutionary theory extends the time horizon of analyses beyond decades and even centuries, which is consistent with the objective of sustainable development. The need for a distant time horizon is especially relevant for research on climate change and biodiversity loss, as these are bound to significantly affect both natural and cultural-economic evolution.

Not surprisingly, climate change research is one of the few areas where optimal growth models have been actually 'applied' (Nordhaus 1994), leading to considerable criticism (for example Demeritt and Rothman 1999, Azar 1998). The issues of uncertainty and irreversibility have been addressed in the traditional economic growth theory context by Kohlstadt (1994), by adding stochastic elements to Nordhaus's (1994) 'DICE' model. The insight obtained is that economic irreversibility due to over-investment in greenhouse gases (GHG) abatement techniques may be more worrisome than the irreversibility of natural processes like GHG accumulation in the atmosphere, climate change and ecological impacts. This is understandable, given the focus on economic efficiency of economic growth in a very simple, aggregate economy, and the neglect of uncertainty associated with the environmental and economic evolution of a complex system (van den Bergh 2004).

Ehrlich et al. (1999) have scrutinized the conventional growth-optimist view that an increase of knowledge and technological change will resolve environmental problems almost automatically. Horgan (1996) claims that the rate of important scientific discoveries is decreasing, while many others have presented an opposite, optimistic perspective. In order to begin resolving this difference of opinion, a distinction can be made between information, understanding and applicable insights. Currently, more researchers than ever before in human history are generating fundamental and applied research. Accordingly, there is more communication of important innovations, notably through academic journals. Ehrlich et al. argue, however, that there is also much disinformation, that is, information that is incorrect, inaccurate or does not add new insights. Moreover, they point out the loss of information due to the destruction of biological and cultural biodiversity.

The literature on economic growth from an evolutionary perspective is consistent with the notion of evolution in that it is very heterogeneous. The basic idea is that of differential growth, which can be seen as a change in the frequencies of all possible individual characteristics. Nelson and Winter (1982, part IV, in ch. 9) developed the first formal evolutionary model of economic growth. Changes in the state of a sector follow probability rules, modelled as a Markov process with time-dependent probabilities, which depend on search behaviour, imitation, investments, entry and selection. If firms make sufficient profits, then they do not search or imitate others; otherwise they do. An important element of evolutionary growth theory is that there is no such thing as an aggregate production function (van den Bergh and Gowdy 2002). Instead, a micro-approach is adopted in which – in principle – individual firms are described. In effect, evolutionary theories propose to avoid an aggregate production function and instead describe diversity of production relationships at the level of individual firms. This idea is also recognized by the neo-Austrian approach, which is formalized using an activity analysis type of model (Faber and Proops 1990). More recently, other formal evolutionary models of growth have been proposed (Conlisk 1989, Silverberg et al. 1988, Silverberg and Verspagen 1994). An important element of evolutionary growth models is that new capital follows from profit being redistributed so that relatively profitable types of capital accumulate relatively fast. This can be regarded as selection such that a technique with a relatively high fitness spreads quickly. This is often modelled by replicator dynamics. In order to complete the evolutionary dimension of the model, selection is complemented by a mechanism of innovation, usually modelled as a stochastic variable in the spirit of the general model of innovation and imitation by Iwai (1984).

A few studies have pursued evolutionary modelling in the area of environmental economics. The famous 'Sugarscape' model by Epstein and Axtell (1996) incorporates a renewable resource (sugar), and thus presents a complex system analysis with evolution and environmental resources. Janssen (1998) and Janssen and de Vries (1998) have incorporated evolutionary elements in climate modelling by allowing adaptive agents to change their behavioural strategies, responding to persistent surprises in global climate as represented by the global mean temperature of the atmosphere. These perspectives include hierarchist – complete control orientation – individualist – adaptive management orientation – and egalitarian – preventive management orientation. The distribution of these perspectives in the population of agents (voters in a democracy) is changed according to a selection process modelled as a replicator equation based on an agent's fitness. This is a function of the difference or gap between expected temperature change and actual temperature change. In other words, when persistent surprises have occurred that cannot be made consistent with the initial perspective on the climate change system and problem, one's perspective adapts. This approach therefore tries to address the lack of complete and correct understanding of climate issues.

Faber and Proops (1990) propose a neo-Austrian approach with evolutionary elements to emphasize the role of time. They allow for the irreversibility of changes in the sector structure of the economy, for uncertainty and novelty, and for a teleological sequence of production activities (roundaboutness). The long-term relation between environment, technology and development is then characterized by three elements:

- The use of non-renewable natural resources is irreversible in time, so that a technology based on it must ultimately cease to be viable.
- Inventions and subsequent innovations lead to both more efficient use of resources and to the substitution of one resource by another.

- Innovation requires that a certain stock of capital goods with certain characteristics be built up.

Faber and Proops construct a multisector model with the production side formulated in terms of activity analysis, which allows them to study the effect of innovation on a transition from simple to more complex or roundabout production activities. Roundabout activities use multiple technologies. For instance, food production has become more roundabout, moving from agriculture with labour through agriculture with labour and capital to a large food-processing industry with many intermediate deliveries. This approach is extended with the technology effects of resource scarcity as indicated above. It can then simulate economic and environmental history from a pre-industrial agricultural society to an industrial society using fossil fuels and capital. A few other studies have employed non-linear models of growth with environment and resources that generate similar patterns (Allen 1997, Clark et al. 1995, Day and Walter 1989). Not all of these, however, are evolutionary in a strict sense, meaning that they are built upon variation, selection and innovation.

Apart from trying to understand and predict growth with evolutionary models, one can raise the first question in the growth debate, namely whether growth means progress from an evolutionary perspective. A problem in answering this question is that progress is a subjective, socially constructed concept that can be defined in many ways. This is reflected by the large number of criteria proposed for identifying evolutionary progress (see Gowdy 1994, ch. 8, Gould 1988, Holland 1998):

- *Increasing diversity*: As diversity holds the key to genetic and therefore phenotypic adaptation to changing circumstances, it is often considered to reflect evolutionary potential. Economic diversity might foster the adaptive capacity of humans and economies in the face of environmental changes (climate).
- *Increasing complexity*: This can apply to morphological structure, or to the number of components or functions. Potts (2000) generalizes this in terms of an increase in the number of connections as well as the levels of nesting of such connections.
- *Extended division of labour*: This can be observed within both natural and social-economic evolution: for example, specialized cells in multicellar organisms, organs, ants in colonies, social organization, complex production activities. This was recognized by both Darwin and Spencer, and provides a link with Adam Smith.
- *New ways of transmitting information*: Living and non-living complex dynamic systems depend on information exchange. This includes

molecular interactions within and between cells, nerves, blood vessels, the lymph system, and chemical substances like hormones, senses (sound, smell, vision), and social interaction through symbols and language. In the economy communication has gone also through various phases: walking, horse, carriage, ship, train, car and plane, and telegraph, phone, fax, and email and Internet. An important aspect has been the continuous growth of the communicating population, which often has been interpreted as 'the world getting smaller'.
- *Population growth*: From an evolutionary perspective, a species is successful if it dominates competitors, meaning dominance in ecosystems and control of its direct environment. In addition, as mentioned above, emergent properties of systems – that is, new levels of reality – seem to be closely related to the size of groups. In other words, a minimum size is needed for a new level to arise. The best-known example outside economics is probably the relationship between dependence of consciousness on brain size or the number of neurons.

From an environmental perspective, one additional criterion should not be left unmentioned. This is the increasing efficiency of energy capture and transformation. In both economic and biological systems evolution can be related to energy processes such as the creation, transmission and stocking of energy (or more correctly 'exergy', which is a non-conserved entity). This has received the most systematic attention in the recent literature on self-organization and complexity (O'Connor 1991, Buenstorf 2000). From an ecological-evolutionary perspective, a rise in energy efficiency means less scarcity and less selection pressure, thus creating opportunities for population growth. The essence here is that evolutionary systems that can tap a larger source of external exergy are able to do more work, and possibly maintain a more complex structure based on internal relationships, each of which requires and uses available energy (exergy). This is clearly illustrated by the immense impact the large-scale access to fossil fuel resources has had on economic activity and growth following the Industrial Revolution. In this context, a theory proposed by Schneider and Kay (1994) is relevant. It states that life (ontogeny) and evolution (phylogeny) involve processes that are manifestations of the second law of thermodynamics. Open systems tend to destroy energy (exergy or entropy) gradients through energy degradation (reducing exergy value) and dissipation (exported or transferred exergy). Systems evolve into more complex arrangements so as to improve degradation and dissipation, and thus destroy energy gradients. This phenomenon finds empirical support in the fact that more complex or developed ecosystems – in the succession towards a climax – tend to be 'cooler';

that is, their radiation has a lower temperature. This means that they are better degraders of energy than less developed ecosystems. This recognition helps identify illegal waste dumps in natural areas using remote sensing techniques. The improved ecosystem capacity for energy degradation goes along with more energy capture, more cycling of energy and material, a more complex trophic structure, more biomass, and more diversity of species. All these elements fit the idea that pathways are created to funnel, recycle and increase the residence of energy (and materials) so that its exergy can be maximally degraded. Schneider and Kay illustrate these ideas with a comparison of 'stressed' and 'control' marshes. Stress leads to less respiration and a less complex food web, that is, fewer channels for energy degradation. The destruction of ecosystems, especially when we consider tropical deforestation, could thus contribute to the warming of the atmosphere.

When thinking about evolutionary progress in economic systems, one cannot ignore the insights of evolutionary biology. According to Maynard Smith and Szathmáry (1995, p. 4), 'The notion of progress has a bad name among evolutionary biologists.' This is reflected in the fact that many professional biologists avoid the discussion of progress or even the term 'progress' in their professional writings, though many of them are advocates of progress in their popular science writings. A well-known expression to illustrate evolutionary progress is the famous phrase 'survival of the fittest' coined by Herbert Spencer, and often incorrectly attributed to Darwin. The correct phrase, however, is 'survival of the fitter', as only some of the unfit disappear, due to coincidence, uncertainty and local search.

Maynard Smith and Szathmáry (1995) suggest that the history of life is better depicted as a branching tree rather than progress on a linear scale. Many species have changed little over the course of hundreds of millions of years. Does complexity increase over time due to evolution? Is there an 'evolutionary hand' that selects and orders diversity, somewhat similar to the 'invisible hand' of Adam Smith? The dominant current view is probably best summarized as evolution being a blind process without any goal in the future. This can increase complexity, but is unable to do so on a continuous basis. There are several reasons why evolution does not necessarily give rise to progress (extending Campbell 1996, p. 433):

1. Selection is a local search process, which leads at best to a local optimum. This can be compared to local optima in complex non-linear optimization problems, studied in operations research. In fact, evolutionary methods are applied to solve such problems and, like the conventional, basic Newton–Raphson gradient type of algorithms, do not guarantee that a global optimum will be found. Maynard Smith and

Szathmáry (1995, p. 5) refer to Fisher's 'fundamental theorem of natural selection', which states that fitness will always increase as a result of selection. However, Fisher's theorem is not general; it requires that the relative fitness of genotypes be constant and independent of the frequency in the population.

2. Organisms are locked into historical constraints. In biology this has been referred to as 'bauplan limits' or 'development constraints' (Gould and Lewontin 1979) or phylogenetic inertia (Wilson 1975, p. 20). In economics it is treated under the headings of increasing returns to scale, path-dependence or lock-in (Arthur 1989). In organization theory it has been called 'structural inertia' and 'imprinting'(Hannan and Freeman 1989, pp. 70 and 205, respectively).
3. Adaptations are often compromises between different objectives, stimulated by a multitude of selection forces. This suggests that evolution is better regarded as a multi-criteria evaluation than as a single-objective optimization.
4. Not all evolution is adaptive micro-evolution. Randomness, genetic (molecular) drift and coincidental founder effects play an important role. In addition, macro-evolution creates boundary conditions for adaptation and may even destroy the outcomes of micro-evolution, in such a way as to set back time ('initialize').
5. Selection can only 'capture' variations that exist. The process of creating variation is limited and largely random.
6. Agents explore only a minor range or subset of the opportunity space, which is reflected in the notion of bounded rationality.
7. Co-evolution means that the environment is not constant and exogenous to the individual species' evolution, but influenced by it. Co-evolution can be regarded as adaptation to an adaptive environment. All straightforward notions of static or dynamic optimization are therefore lost, since the boundary conditions in a constrained optimization formulation of evolution are not even known then. Co-evolution also has historical, path-dependent features. It can be cast in the adaptive landscape metaphor (Kauffman 1993), in which case the landscape becomes something like a 'choppy sea rather than something forged in granite. In which case there would be no real progress, for as soon as one climbed a peak it would collapse beneath one' (Ruse 1999, p. 118).

Notions of progress, happiness and equality are more typical of the social sciences than of biology, and play a core role in economics. Archaeology and anthropology have used historical classifications that reflect notions of progressive and stage-like change – think of the Stone, Bronze and Iron Ages, or formative, classical and post-classical stages, or group size terms such as band, tribe, chiefdom and state. Early writings on social evolution employ terms such as savages and primitive civilizations, which reflect similar connotations. A correct interpretation of evolution suggests strongly that the old economic idea that markets can realize a social optimum, the invisible hand of Adam Smith, is incorrect. Sen (1993) further notes that evolution as improving species does not imply improving the welfare or quality of life of each individual organism. Fitness is not a useful criterion for progress in general, even if the evolution of a species into one more efficient in food gathering, moving faster (running, swimming, flying), or more effective in performing certain functions, can usually be regarded in terms of increased fitness. The reason is simply that a higher fitness and survival do not necessarily mean a happier or more pleasant life.

A problem for the social sciences is that welfare above a certain threshold, as defined by (basic) needs, is relative. Economic growth without significant changes in equality does not change the relative welfare of individuals, and is thus not very useful. Evolution as a continuous change in diversity, however, implies that inequality will arise again and again. Distributional change and inequality are not exceptional and possibly inherent to evolution: firms in terms of size and (market) power, consumers and households in terms of income and wealth, social groups in terms of social and political power, and countries in terms of all these variables. In this respect, it may be noted that the repeated selection for fitness implies that populations and species would be continually stimulated to improve their fitness, since otherwise they would be overtaken by others. This is known as the 'arms race' or 'Red Queen hypothesis' (Strickberger 1996, p. 511). Note that this suggests that economic growth is perhaps a natural phenomenon, not one that will necessarily lead to a continuous improvement in human conditions, but one that is inevitable given the continuous change created by strong evolutionary forces in the modern economy. This is opposed to a view that human societies such as isolated tribes, or even societies during the Middle Ages, used to live in a stationary state – negatively referred to as 'stagnation' – in static balance with their local environment. Modern Western societies, notably since the Industrial Revolution, are instead on an unsustainable and non-equilibrium growth path.

4. THE ROLE OF ENVIRONMENT AND RESOURCES IN LONG-TERM ECONOMIC HISTORY

In economics, growth is usually isolated from biological–social reality and even technological reality. Moreover, it is considered a-historical, describing an extremely aggregate process driven by reversible mechanisms. But history is full of irreversible changes resulting from the enormous diversity of agents and techniques in the economy. With its emphasis on variable diversity, evolutionary growth theory is a step in the direction of historical analysis. This section will argue that in order to create an accurate historical context for economic growth, one needs to include both evolutionary mechanisms, such as variety, selection and innovation, and environmental and resource dimensions. This section largely summarizes ideas first proposed in van den Bergh (2002).

Economic evolutionary history can start as far back as the hunter–gatherer societies. The environmental impact of these was negligible due to the small disparate groups of humans living in largely natural ecosystems. The negative effects were mainly local and restricted to land cover altered by fires caused by humans. The next phase is the transition to agriculture, which is hypothesized to have been due to a combination of climate change and sufficient variation of flora and fauna in ancient Mesopotamia. Social–cultural development accelerated after the last ice age (about 13 000 years ago) because of the development of sedimentary agriculture, which allowed for labour division and specialization. Agriculture meant the settlement and concentration of people in certain areas, which gave rise to human alterations to the environment. The third major phase was set in motion by the Industrial Revolution. This has been hypothesized to have been stimulated by firewood scarcity in England. The structural change known as the Industrial Revolution was made possible by the critical mass of large communicating 'populations' of scientists and pragmatic inventor–engineers in England at the time. The enormous growth since then, involving both population growth and economic growth, led ultimately to an environmental crisis during the second half of the twentieth century. From an environmental perspective, the twentieth century is exceptional in human and natural history. This is illustrated by a large number of indicators. During that century world population quadrupled, the global economy expanded 14-fold, energy use increased 16 times, and the 'control' of world biomass increased to 40 per cent. From an evolutionary perspective one can see the extreme and rapid change of the environment as a clear case of a species, in this case *homo sapiens*, being maladapted.

Few complete models are available to study and understand immense changes like those described. Several authors have tried to come up with evolutionary frameworks (for example Gowdy 1994, Mulder and van den Bergh 2001). These have been argued to be consistent with the theory of punctuated equilibrium (Somit and Peterson 1989, Mokyr 1990a and 1990b, Gowdy 1994), although so far this is no more than a loose conceptual connection. A very general type of model is based on the notion of co-evolution. This was originally proposed in ecology to refer to the joint evolution of butterflies and flowering plants (Ehrlich and Raven 1964). Indeed, co-evolution integrates elements from evolutionary biology and ecology to denote the evolution of species interactions. Currently, however, co-evolution is used in a much wider sense. It has been used to refer to the following kind of interactions: biological–cultural, ecological–economic, production–consumption, technology–preferences, and human genetic–cultural (Lumsden and Wilson 1981, Norgaard 1984 and 1994, Durham 1991, Gowdy 1994, and Wilson 1998). An interesting typology of co-evolution comes from Durham (1991):

1. *Genetic mediation (interactive mode)*: Genetic changes affect cultural evolution.
2. *Cultural mediation (interactive mode)*: Cultural changes affect genetic evolution.
3. *Enhancement (comparative mode)*: Cultural change reinforces natural evolution.
4. *Opposition (comparative mode)*: Cultural change goes against natural evolution.
5. *Neutrality (comparative mode)*: Cultural change is independent of biological evolution or selection.

Hinterberger (1994) even tries to link co-evolution to Dawkins's (1976) notion of 'memes' (see also Aunger 2000). Although co-evolution focuses on genetic–cultural interactions, it can serve as the basis for a generalization of other types of co-evolution, such as the interaction between evolutionary economic and ecological systems. Note that co-evolution is, however, often used in a loose manner, without including aspects of populations and diversity. In this case, the term 'interactions' of subsystems would be a better term than co-evolution.

Environmental factors may have influenced crucial changes during the cultural–social history of humankind. Potential environmental factors of influence are local and global climate, the diversity of potential food resources in the form of available local plants and animals with sufficient concentrations of proteins, carbohydrates, fats and vitamins, unfavourable soil conditions (erosion due to irrigation in Mesopotamia), and the scarcity of fuels (notably fuelwood). Diamond (1997) summarizes the large literature that has tried to support the theory that the availability of animal and plant species stimulated early domestication and thus agriculture and settlements. He notes that the sufficient diversity of agricultural experimentation was only possible in continents with the major axis being east–west oriented (the prime example being the Eurasian continent), as this allows for the spread of agricultural technologies among regions with similar climates. Diamond considers this as an important reason for the early economic success of Eurasia. His theory thus explicitly relates early economic development to geographical and resource factors. His ideas are surprisingly close to notions of spatial diffusion of technology, which are so common in neo-Schumpeterian theories of technical change.

Wilkinson (1973) has developed an ecological theory of economic development that aims to link the Industrial Revolution to natural resource factors (see also Common 1988). It recognizes a number of human strategies to respond to resource scarcity, such as new techniques, new resources, new goods and migration. Wilkinson's ideas imply an environmental perspective on the origins of the Industrial Revolution at the end of the eighteenth century. This started with agriculture and iron smelting, which used large amounts of timber for energy purposes, in turn giving rise to a significant loss of forest cover in England. The resulting shortage and high price of wood stimulated the use of coal. In the early phase, this focused on opencast coal mining. Later deep mines were explored. This created an important problem, namely that groundwater needed to be pumped out, and this led to the first large-scale application of the steam engine. In turn, the widespread use of the steam engine gave rise to various refinements and competing alternative models. In a next phase, spin-offs to other sectors occurred, especially to the textile industry and to transport, through ships and trains powered by steam locomotives.

In addition, a complete view of macro-history since the Industrial Revolution not only involves growth trends but also cycles or long waves. These are caused by major shifts in technology, due to fundamental advances in science. A rough classification of waves since the Industrial Revolution is shown in Table 11.3. Long waves have gone along with a number of changes. Among other things, the average size of firms has increased, the research (R&D) and innovation process has shifted from firm to international levels, firm interactions and industry structure have changed, and new key factors and sectors have arisen. In addition, each period has had its own environmental impact, as illustrated in the table.

An important dimension of long-run historical cultural–economic evolution has been the size and composition of basic human groups. According to McNeill (2000), at the time of the transition to agriculture, around 8000 BC, humans were still outnumbered by other primates, notably baboons. The current human population is about 100000 times as large as that of other species with a similar body size and position in the foodweb. Much of historical growth went along with an increase in the size of human groups, ultimately leading to cities and states, a process that has not yet stopped (in a large part of the developing world). Much of economic (GDP or GNP) growth is due to population growth. Yet standard

Table 11.3 Environmental and resource aspects of long waves

Phase	Core resource	Environmental impact
Early Industrial Revolution	Coal	Urban pollution
Steam power and railway	Coal	Large-scale infrastructure
Belle époque	Coal, steel	Heavy industry related
Fordist mass production	Oil, synthetics, heavy metals, fertilizers	Car related (noise, exhaust pollution, road infrastructure), toxic substances, acid rain, soil erosion
Information and communication	Heavy metals, tropical wood	Biodiversity loss, global warming
Future	Genetic resources, water?	Genetic pollution, climate change?

Source: van den Bergh (2002).

economic growth theory has devoted only little attention to population issues, even if these are evidently of great importance to economic pressure on natural resources and the environment.

The size of human groups might even have been crucial for certain major transitions during human–economic evolutionary history. For instance, the Industrial Revolution might not have occurred so early had it not been for large populations of engineers and academics. The Netherlands, for example, probably did not exceed the required 'critical mass'. Compare this to the critical number of brain cells before certain higher-level phenomena emerge such as memory and consciousness. Other examples are the number of cells in a body before certain hormone processes are triggered, and the number of ants in a colony before certain physical structures can be developed.

Diamond (1997, ch. 14) identifies four main stages of social structure that humankind can be considered to have moved through in the last 13000 years. Through history, non-kin relationships have increased in importance. Family/band (5 to 80 people), tribe (hundreds of people), chiefdom (thousands of people), state (more than 50000 people). This increase in group size went along with a number of other structural changes. The major ones are social (hierarchical) organization, formal conflict resolution, labour division, specialization, and trade and transport. A globally intensively interacting human population – due to 'globalization' – could be the last stage. Which emergent properties this might have cannot by definition be known beforehand.

5. ECOSYSTEMS AND RESOURCE MANAGEMENT

At the level of ecosystem and resource management, various issues are involved in relation to evolutionary economics. The dynamics of ecosystems involves reversible dynamics such as population growth and ecosystem succession, and irreversible changes due to selection, as well as more fundamental evolutionary changes. Ecological economics is very much influenced by the notion of resilience, which is an extended stability concept (see Figure 11.1). Perrings (1998) mentions two alternative readings: one is directed at the time necessary for a disturbed system to return to its original state (Pimm 1984); the other is directed at the intensity of disturbance that a system can absorb before moving to another state (Holling 1973). In line with the latter interpretation, resilience has been called 'Holling sustainability', as opposed to weak 'Solow–Hartwick sustainability' (Common and Perrings 1992, Ayres et al. 2001). As a result, ecological economics studies pay much more attention to the sensitivity of ecosystems at the micro-level, whereas traditional environmental extends economic growth theory with environmental variables, emphasizing determinism and coarse long-term trends in a macro-approach that lacks any micro-detail. This critique sounds similar to the one by evolutionary economics of neoclassical (exogenous or endogenous) growth theory: in simple words, too crude and aggregate. Resilience is often linked to biodiversity, as more diversity enhances stability of systems. However, this is still being debated. At the moment resilience is even examined as an analogy for the functioning of social systems such as bureaucracy, politics and economy (see Levin et al. 1998). Holling (1986) himself has proposed a four-box model. It depicts ecosystems and their changes in a two-dimensional diagram with axes labelled 'potential'

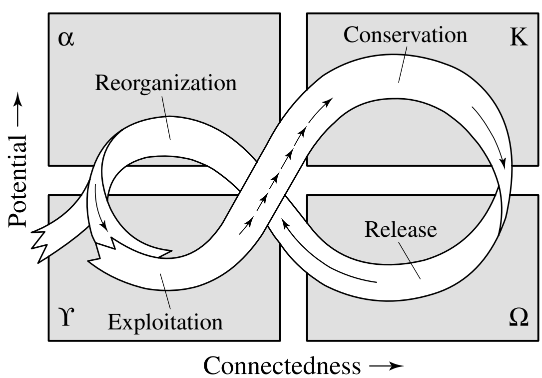


Figure 11.1 Resilience according to Holling (1986).

(biomass) and 'connectedness' (complexity of the foodweb). Ecosystems can then go through the four phases: 'exploitation', 'conservation', 'release' and 'reorganization'. The 'release' phase, for instance, is set in motion by forest fires, storms and outbreak of diseases. In the four-box model management aimed at artificially prolonging a certain phase, notably conservation, can reduce the resilience of the system. For example, checking small forest fires, which leave seeds intact, will result in an accumulation of forest biomass. This in turn will increase the probability of the occurrence of a large forest fire, going along with very high temperatures, which in turn can destroy plant seeds and thus prevent the reorganization phase from occurring successfully. Note, finally, that some authors have tried to find analogies for resilience in socio-economic systems (see Levin et al. 1998).

An important type of impact of human resource use on the ecosystems and populations of living organisms is the selective pressure exerted by managing and harvesting practices. The general nature of the problem is that resource harvesting affects not only the quantity of the resource but also its quality, or its composition in genetic terms. Examples can be found in agriculture (use of pesticides and herbicides, monocultures), fisheries (mesh size, season of fishing), ecosystem management (groundwater level, fire protection), and health care (antibiotic use leading to selection of resistant strains of bacteria). These are all special cases of biodiversity loss. A number of economic and management issues are at play (Munro 1997):

- Since the rate of mutation is low, selection is the main mechanism to be addressed (resistance is pre-adaptive/non-Lamarckian). The question then is what impact humans have on selection.
- Resistance usually has a (small) cost, otherwise the resistant individuals would be well represented even without selection pressure from pesticides and the like. In other words, the resistant individual is more susceptible to other stress or selection factors. Withdrawal of the selection pressure will then usually lead to a decline in the proportion of resistant individuals in the population. Nevertheless, selection pressure will usually lead to irreversibly reduced (bio)diversity in the population. This is the essence of the problem, creating a persistent externality towards the future. Moreover, the renewed introduction of the selection pressure will very quickly achieve widespread resistance due to the relatively large proportion of resistance-affecting genes in the population.
- Is there evidence of resistance to one type of insecticide creating resistance to others? If so, then a strategy of shifting between different insecticides will not be very effective. This seems to depend on the type of (metabolic/biological/chemical) mechanism through which

individuals achieve resistance. This can include the excretion of the unabsorbed insecticide, changes in behaviour (reduced resting behaviour indoors in rooms sprayed with insecticide), or the reduced penetration of the chemical in the body. Empirical data show that the sequence of rotation of applying different insecticides as well as specific locational factors such as climate can make a difference.

- Are the costs of overcoming resistance increasing? In spite of learning effects such as the development of new antibiotics and in part due to cross-resistance, the costs have increased, often significantly.
- Does the above provide sufficient reasons for policy intervention? This depends on how much one is concerned with future effects, on present or future generations. Lack of adequate policy implies myopic decisions by economic agents, which do not take into account future external costs. This can be considered the traditional environmental economics perspective (which may also include the idea that property rights are missing). From an evolutionary angle, the irreversibility or even irrevocability of changes is worrisome, especially when put in a wider ecosystem context, which is only partially understood: empirical predictions about the impact on other species are very uncertain. On a greater time scale, evolutionary consequences are even more uncertain, including evolution which goes beyond pure selection, such as sexual recombination, mutations and the disappearance or emergence of (new) species.

Munro implements these assumptions by extending the standard models of optimal use of a renewable resource and of optimal pesticide use with a dynamic negative externality with a description of genetic selection. He focuses on the example of the use of insecticides that raises the fitness of resistant insects relative to susceptible competitors. The perspective can be that of an individual farmer or more realistically a policy-maker trying to understand the optimal dosage of insecticide.

The main insights obtained by analytically solving the resulting optimization problem are as follows. First, given initial conditions stating that the pest is at its carrying capacity level and the proportion of resistant individuals is extremely small, then a fixed dosage that makes resistant pests fitter than susceptible ones leads to a pattern in which initially both the total pest population and the proportion of susceptible individuals decrease. Once the proportion of resistant individuals reaches a critical point, the population as a whole recovers and returns to its original maximum (carrying capacity) level. Evidently, this is unattractive from an economic benefits perspective. The initially effective policy in time becomes completely ineffective.

Optimization of the net present value of agriculture without taking into account the selection impact of pesticide use reflects myopic decision-making, because it is based on the assumption that the proportion of susceptible individuals is fixed (that is, an unchanging genetic composition of the population). This can be regarded as the traditional planning that does not anticipate evolution. Instead, optimization under perfect foresight about evolutionary consequences of insecticide use gives the fully optimal plan. Comparison of the myopic and fully optimal plans shows that the optimal value of the size of the pest population is lower, while the proportion of susceptible individuals is higher. Pesticide use remains positive, with a level related to the gap between the fitness of susceptible and resistant organisms. If the discount rate becomes sufficiently large, then optimal pesticide use will be higher so that resistance will increase. As a corollary, R&D on new pesticides will be too high under myopia.

Van der Voet et al. (2000) analyse the evolutionary consequences of human dominance in the biosphere at a more general level. They argue that impacts are very heterogeneous for different types of living organisms. The three most influential factors have been the occupation of space through its influence on habitats, the harvesting of biotic resources, and the deliberate and unconscious introduction of exotic species in ecosystems (Perrings et al. 2001). Other factors are the influence on local and global nutrient cycles (direct influence on the growth rates of plants), the emission of toxic substances (the influence on predators far up in food chains), the fragmentation of populations through physical infrastructure (roads, and rail- and waterways), the domestication of animals and plants, and traditional and modern biotechnology (breeding and artificial selection, using living organisms in production, and the use of genetic technologies to create genetically modified organisms). In evolutionary terms one can classify influences as changing the selection environment (predation, nutrients, habitats, chemical pollutants), introducing new mutations (exotic species, chemical pollutants), affecting spatial isolation (infrastructure), and influencing reproductive advantage (pesticides). The ultimate consequences are extinction or an influence on speciation, thus affecting the direction of natural evolution. Predators and prey species are especially sensitive to this, for different reasons (van der Voet et al. 2000). Predators live in small numbers, follow an ecological K strategy, require a large habitat, and often have no or a relatively low (or even negative) economic value. They are especially sensitive to land use, toxic substances (concentration in the food chain) and hunting. Due to a low reproduction rate they are not very adaptable. Prey have a positive economic value, and remain part of a pre-agricultural hunting culture. Harvesting pressure is the most serious threat, especially for fish populations. Due to mesh size and the season of harvesting, changes occur in the size and age distributions of harvested populations. Most prey species predated by humans have been domesticated; the last major group of 'free' prey species, that is, fish populations, are now being increasingly domesticated as well (fish farms). This might be considered as a sort of final stage in cultural–natural co-evolution. Pests, including exotic species and escaped pets, generally have a very negative economic value, either through disruption of agricultural productivity, competition with endemic species, or the costs of pest-treatment (pesticides). They are r-strategy species that reproduce very rapidly, allowing for quick adaptation and resistance to pesticides. Current trends suggest that predators and prey species lose ground, pets are little affected, and pests gain. It is very difficult to say what global consequences this will have, as both the loss of biodiversity and the ecological impacts are still little understood.

Much attention in environmental economics has been given to the risk of over-exploitation of common property or common-pool resources, such as fisheries (Hardin 1968, Bromley 1991, Ostrom 1990). Although common property is often confused with open access, where exploitation is more likely, in common property resources the risk is also serious. It depends very much on the type of common property regime that is active, and may therefore differ from situation to situation. An important question is whether resource conflicts and over-use should be responded to with strict policies set by higher-level governments, or whether one should rely on the endogenous formation of use regimes. An evolutionary perspective has been used to analyse the latter, based on the idea that such regimes need to be sufficiently supported by the individuals participating, or that a single norm will evolve (Axelrod 1986). One finding is that externally imposed rules and monitoring can reduce, destabilize or even completely destroy cooperation (Ostrom, 1990). When monitoring is imperfect, results are even worse, and stimulating norms through communication is certainly more desirable than external regulation. The latter is only desirable if monitoring and sanctioning can be well developed. Yet this self-organization process in the most fundamental and general form is probably still not entirely understood, for instance the influence of group size. Instability in the evolutionary equilibrium can arise when certain parameters change, such as the resource price, and interestingly also when other rules (state property, open access) are implemented by an external regulator. In the latter case norms may erode, ultimately leading to resource extinction. Equilibrium can also break down when sanctions decline, harvesting technology becomes more productive, or the price of the resource increases. Other risks to evolved norms are migration, external economic and political disturbances, and natural disasters. These issues have been examined using a wide range of empirical approaches, such as empirical field studies (Martin 1989/1992) and laboratory experiments (Ledyard 1995).

Studies combining economic evolution and renewable resource dynamics are rare. Noailly et al. (2003a) present an analytical analysis of the basic problem. They step away from the traditional assumption of rational harvesters and assume instead a distribution of harvesting strategies that changes according to replicator dynamics. The interaction of evolutionary and resource dynamics leads to the result that both strategies may survive in the long run. Another study in this vein has started more explicitly from evolutionary game theory (Sethi and Somanathan 1996). It models changing norms in the context of competing users of a commonly owned renewable resource, such as pastures (grazing of cattle, herding), lakes and seas (fisheries). The model fits into the evolutionary game theory tradition. With a fixed number of resource users a continuous range of Nash equilibrium exists, creating a choice problem. This problem is solved by letting the evolutionary dynamics of interaction of resource use and social interaction strategies 'select' the 'feasible equilibrium'. This model is the first to combine two dynamic relationships: evolutionary (replicator) dynamics and renewable resource dynamics. The evolutionary model is set up as a two-stage game: (1) resource use strategy, (2) punish or not. Punishment is assumed to be costly. There are three possible strategies: defecting (high effort); enforcing (low effort, sanctioning of defectors); and cooperating (low effort, no sanctioning of defectors). The profits under each strategy are variable and depend on the distribution of strategies. The dynamics according to selection of highest pay-off are modelled as replicator dynamics. In summary, there are really only pure equilibria: defector or restraint (cooperative/enforcers). In the end nobody or everyone cooperates. Note, however, that this is not the case in spatial interaction games, in which the spatial structure of agents, their interactions and the resource are taken into account (Noailly et al. 2003b). The game is a combination of prisoner dilemma and coordination games. Once a defector equilibrium is reached, recovery of cooperative behaviour is difficult.

6. TOWARDS AN EVOLUTIONARY THEORY OF ENVIRONMENTAL POLICY

Environmental policy theory is completely based on applying neoclassical welfare theory. It links a competitive equilibrium to a situation with maximum social welfare or, at least, Pareto efficiency. However, evolutionary systems based on bounded rationality, non-equilibrium and path-dependence imply that the normative part of neoclassical economics is based on incorrect assumptions. If the correspondence between market equilibrium and the social welfare optimum is lost, then the two fundamental theorems of welfare economics are lost. As a result, planning or market solutions cannot be guaranteed to deliver socially optimal outcomes. Since evolutionary economics cannot offer a similar normative perspective, its major general policy implication is negative.

Nevertheless, alternative policy suggestions can be derived, namely by considering in turn the elements bounded rationality and path-dependence. Bounded rationality in an evolutionary context means that individuals often show automatic or habitual behaviour, and that they tend to imitate others, especially in complex and uncertain situations where information-gathering is costly. These different alternatives imply a deviation from traditional policy insights. For more details, see the illustrations of environmental policy in Van den Bergh et al. (2000). A summary of their findings is as follows. Habitual behaviour is one explanation for the 'energy gap' or the unreaped economic benefits associated with potential energy conservation measures inside many firms. Behaviour according to lexicographic or hierarchical preferences means that the substitution of regulated commodities is not always possible. This undercuts the empirical fact that growth of material consumption, a higher need, has occurred at the cost of some lower 'immaterial' needs, such as rest, the absence of stress, personal contacts, social life, serenity (no noise), no violence and so forth. As opposed to neoclassical preferences – giving rise to continuous and differentiable utility functions – lexicographic preferences also suggest that individuals are unable to make a continuous trade-off between environmental functions or resources and economic goods. This can explain why many individuals are unwilling to respond favorably to questions about willingness to pay for environmental changes (Spash and Hanley 1995, Stern 1997). Therefore, lexicographic preferences can be related to the notion of strong sustainability, where complementarity, uniqueness and non-substitutability of life support functions, climate regulation and nature in general are emphasized. With regard to uncertainty, individuals often employ simple rules to limit the amount of information they need to consider to make a decision. Prospect theory and regret theory stress the asymmetry with which individuals perceive and deal with gains and losses (Kahneman and Tversky 1979). This may affect how individuals interact: Coasian negotiations between polluters and victims, negotiations between countries about environmental agreements, or negotiations between regulators and polluters about the initial distribution of permits. Roe (1996) applies 'Girardian economics', which starts from pervasive uncertainty. Mimetic behaviour is seen as resulting from it, leading to reduced economic diversity of behaviour, strategies, activities, ideas and products. This model is applied to the problem of sustainable development, which leads to the following conclusions: sustainable development is a social convention which for a time stabilizes decision-making under high uncertainty. The response should be to reduce uncertainty that drives the crises of undifferentiation (homogeneity) due to imitation. Some specific elements of such a strategy are to buffer or decouple resource systems and their management from environmental uncertainty; to encourage diversity via evolution and the diffusion of more than one kind of sustainable development; and to treat sustainable development on a case-by-case basis. The last links both to the discussion on biodiversity, where maximum diversity keeps most options open and maximizes the resilience of any system. The whole approach is very much in the spirit of evolutionary economics. The idea that preferences are endogenous instead of invariant has led Norton et al. (1998) to argue that changing consumers' preferences can be an instrument of environmental policy. In particular they state that stable preferences are at best realistic over short periods of time, and that sovereign preferences are inconsistent with the long-run goals of sustainability. Consequently, public discussion about ethical consumption and sustainability should be stimulated through education, advertising rules and cultural norms. Changing consumers' preferences through democratic processes could be used to encourage environmentally conscious consumption in a way that consumers would not feel 'deprived and unhappy' but 'enlightened and happy after being educated into the joys' (Norton et al., 1998, p. 203). Most democratically elected governments formulate public policies with the aim to influence norms that are considered as criminal, racist or otherwise undemocratic, so why not extend this towards environmental sustainability?

From the perspective of path-dependence and lock-in, a question is how regime shifts occur and how they can be stimulated. This is related to the problem of how to avoid a lock-in of inefficient or undesired technologies. Preventing early lock-in requires portfolio investment. The unlocking of undesired structures and technologies from an environmental or some social welfare perspective cannot be done only by 'correcting prices', but requires a combination of policies. Examples are well known in the area of environmental policy. Setting a clear overall goal like 'zero emission' in California meets the requirement of reducing policy uncertainty. Creating semi-protected niches can stimulate renewable energy sources, notably solar energy based on photovoltaic cells. Other elements of a strategy to avoid lock-in include stimulating the diversity of R&D, pathway technologies (electric batteries) and complementary technologies. Of course, price-based instruments are still important, focusing on cost-effectiveness and 'dynamic efficiency', but are insufficient. Fundamental, university research provides the basis for major technological change. In the environmental area, important changes envisioned now are decentralized energy production based on renewables (solar and wind energy), precision-biological agriculture and genetic technology, low-pressure/temperature chemistry (catalysis), nanotechnology (dematerialization, waste and emission reduction), and battery-powered electric vehicles. In addition, social or organizational innovations may require governmental support, such as car sharing or mixed car–public transport systems. Governments can change the selection environment for car producers, for example, through technical limits on motor size, speed and acceleration power. This can stimulate technological innovations directed at slower and lighter cars. This would have several advantages: less materials and energy used in production; less energy involved in collisions and reduced risks (and externalities) for sensitive road users (pedestrians, cyclists, motorcyclists, and users of relatively small cars). For other issues related to technology and environmental policy from an evolutionary angle, see Kemp (1997).

Finally, evolutionary thinking and modelling have several interesting general implications for public policy. One can test which policies are robust against endogenous evolutionary phenomena. This can be regarded as an extended type of stability analysis. Evolutionary theory suggests a fundamental alternative to policy in that it regards public policy as the outcome of a difficult and fragile process of communication and consensus-building among large groups of human individuals. In many policy contexts the effectiveness of public policy is small because individuals cannot be accurately monitored. In such a case, it may be preferable to allow for the evolution of a set of common norms or rules, based on direct interaction among spatially proximate individuals. This does not necessarily imply a *laissez-faire* strategy, since governments will still play an important role in providing suitable boundary conditions such as legal arrangements. These ideas have been developed in political science and are now coming to economics. They are particularly relevant to the study of the management of environmental resources, as many of these are instances of common-pool resources, which again are a special type of public goods.

7. CONCLUSIONS

The link between evolutionary and environmental economics has many different aspects. The most fundamental one is perhaps that in the long run – both in history and in the future – environmental problems and resource scarcity are important causes of major economic change. This has been argued here to hold in any case for the rise of agriculture and the Industrial Revolution. The environmental and resource factors which underlie economic evolution have been largely neglected by evolutionary economists and economic historians alike, and certainly do not occupy a serious place in standard treatments. For the future, the relationship between economic evolution and environmental resources requires that the notion of sustainable development be built on an evolutionary perspective. This involves the difficult but exciting task of integrating disparate literatures on technology and innovations, co-evolution and environmental history. It will give rise to a mixture of biological, social, technological and economic evolutionary processes, depending on the problem and exact time horizon. For instance, the environmental and health impacts of modern biotechnology based on genetic engineering will surely have consequences for biological, human and economic evolution alike.

The concepts and suitable methods of evolutionary analysis depend very much on the scale on which environmental issues are being studied. This can vary from the ecosystem level to the macroeconomic level. Integration of evolutionary economic models and environmental models results in very complex overall model systems. This is due to the combination of two dynamic relationships, namely an evolutionary one covering selection and innovation processes, and an environmental one involving ecosystem or renewable resource dynamics. Moreover, if natural evolution is given a place next to economic evolution, then two evolutionary systems will interact, leading to all kinds of possible co-evolutionary dynamic patterns. Currently, it seems that a spatial approach addressíng this interaction will be most fruitful.

Over the past decades, economists have put much effort into research on environmental and resource issues. This has given rise to a field known as environmental (and resource) economics, and more recently to a multidisciplinary field known as ecological economics. Although both fields follow the common path of academic research towards more specialized studies, there seems to be a lack of theoretical innovation. Evolutionary economics could provide the necessary theoretical input to be combined with the multidisciplinary approach of ecological economics.

Global environmental problems, notably the enhanced greenhouse effect and the associated risk of climatic change, present us with the most difficult of all environmental problems. Current economic theory, with its focus on welfare optimization and cost–benefit analysis, is insufficiently capable of addressing aspects typical of such problems. The most important one is surely that possible extreme changes in the global environmental conditions for life on this planet will affect economies in unprecedented ways. I think that evolutionary approaches can provide unique perspectives on such issues. They recognize the importance of key elements of dynamic systems, such as diversity, innovation, selection, path-dependence, structural change, emergence, and slow and fast dynamics. These elements are insufficiently valued in current economic research on climate change, but can be fruitfully linked to current practices and insights, notably in environmental and ecological economics. Indeed, heterogeneity has received much attention in microeconomics recently, but it is not yet, as in evolutionary economics, regarded as the driving force behind system dynamics.

In effect, an evolutionary theory of growth, resources and the environment needs to be developed. This can be achieved by combining conceptual, historical, theoretical and applied work. Theory should link different approaches in evolutionary modelling, including evolutionary game theory, multi-agent modelling, network modelling, cellular automata, and innovation–selection–diffusion models. The desired outcome of this is a new class of spatial evolutionary models, which can adequately capture the essence of environmental problems and human responses to these. The spatial dimension is important because both economic and environmental systems are spatially heterogeneous, characterized by spatial diffusion as well as interaction, and show spatial patterns and structures.

From a somewhat different and more general angle, one could say that an important question is how complex systems of interactive behavioural strategies perform and develop in an environment that is inherently dynamic. Considering typical, generalized environmental dynamics can help us to examine the influence of environmental changes, such as trends, seasonal cycles, erratic fluctuations, or extreme events, on the stability of social-economic systems. The generality of this question is clear. Dynamic environments are typical of economic analyses of environmental and resource problems. But environmental dynamics can derive from a number of other situations as well, such as macroeconomic context (business cycles), technological innovations (Internet), changes in policy and regulation (tax system revision), and demographic changes (both population size and structure).

At a policy level, different regimes of self-regulation and external regulation can be compared, based on creating conditions for self-regulation – through property or use rights, moral suasion, direct regulation, or flexible price-based incentives. The relevance of the international level is illustrated by the emergence of international agreements. The positions and resulting interactions among participating countries are influenced by the variation of economic and natural conditions in these countries. These relate to differences in local climates, soil conditions, vegetation cover, demography, economic activities, technical innovation and education levels. Such differences are most extreme between developing countries in tropical arid regions, and industrialized countries in temperate climate zones.

An important question in the growth debate within environmental economics is whether growth is equivalent to progress. It is interesting to consider this from an evolutionary angle. At some high level evolution can be regarded as progress by definition, otherwise it would not have been recognized in the first place – the notion would not exist. Evolution has some elements of directionality and progress. Nevertheless, evolution certainly is not identical with continuous progress. Furthermore, much of what is presented as progress by optimists often possesses mixed blessings. In any case, evolutionary theory suggests that irrespective of one's beliefs, the hope that economic growth will come to a halt at some maximum level of welfare is idle. The reason is simple: the evolutionary mechanisms of economic selection and innovation remain operative, unless something really terrible happens. One relevant scenario that comes to mind is the collapse of modern economies due to severe climate change. But as Georgescu-Roegen (1972, p. 35) has noted, 'Perhaps, the destiny of man is to have a short, but fiery, exciting and extravagant life rather than a long, uneventful and vegetative existence.'

NOTE

* An anonymous referee provided useful comments.

1. These opportunities for mutual learning have been clearly recognized in the Netherlands, where the Netherlands Organization for Scientific Research has just launched an eight-year integrated research programme on 'Evolution and Behaviour' that involves biology, behavioural sciences (psychology) and social sciences.

REFERENCES

Allen, P.M. (1997), 'Evolutionary complex systems and sustainable development', in J.C.J.M. van den Bergh and M.W. Hofkes (eds), *Theory and Implementation of Economic Models for Sustainable Development*, Dordrecht: Kluwer Academic Publishers.

Arrow, K.J., Solow, R., Portney, P., Leamer, E., Radner, R. and Schuman, H. (1993), *Report of NOAA Panel on Contingent Valuation*, 58 Fed. Reg. 46.

Arthur, B. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *Economic Journal*, **99**, 116–31.

Aunger, R. (2000), *Darwinizing Culture: The Status of Memetics as a Science,* Oxford: Oxford University Press.

Axelrod, R. (1986), 'An evolutionary approach to norms', *American Political Science Review*, **80** (4), 222–38.

Ayres, R.U. (1994), *Information, Entropy and Progress: Economics and Evolutionary Change*, Washington, DC: AEP Press.

Ayres, R.U., Bergh, J.C.J.M. van den and Gowdy, J.M. (2001), 'Strong versus weak sustainability: economics, natural sciences and "consilience"', *Environmental Ethics*, **23** (1), 155–68.

Azar, C. (1998), 'Are optimal CO_2 emissions really optimal?', *Environmental and Resource Economics*, **11**, 301–15.
Baumol, W.J. and Oates, W.E. (1988), *The Theory of Environmental Policy*, 2nd edn, Cambridge: Cambridge University Press.
Bergh, J.C.J.M. van den (ed.), (1999), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Bergh, J.C.J.M. van den (2001), 'Ecological economics: themes, approaches, and differences with environmental economics', *Regional Environmental Change*, **3** (1), 13–23.
Bergh, J.C.J.M. van den (2002), 'An evolutionary analysis of growth, environment and resources', in R.U. Ayres, D. Simpson and M. Toman, *Scarcity and Growth in the New Millennium*, Washington, DC: Resources for the Future.
Bergh, J.C.J.M. van den (2004), 'Optimal climate policy is a utopia: from quantitative to qualitative CBA', *Ecological Economics*, (forthcoming).
Bergh, J.C.J.M. van den, A. Ferrer-i-Carbonell and G. Munda (2000), 'Alternative models of individual behaviour and implications for environmental policy', *Ecological Economics*, **32** (1), 43–61.
Bergh, J.C.J.M. van den and Gowdy, J.M. (2000), 'Evolutionary Theories in Environmental and Resource Economics: Approaches and Applications', *Environmental and Resource Economics*, **17** (1), 37–57.
Bergh, J.C.J.M. van den and Gowdy, J.M. (2002), 'The microfoundations of macroeconomics: an evolutionary perspective', *Cambridge Journal of Economics*, **27** (1), 65–84.
Bergh, J.C.J.M. van den and Hofkes, M.W. (eds) (1998), *Theory and Implementation of Economic Models for Sustainable Development*, Dordrecht: Kluwer Academic Publishers.
Bergh, J.C.J.M. van den and de Mooij, R.A. (1996), 'An assessment of the growth debate', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Lyme, USA: Edward Elgar.
Bergh, J.C.J.M. van den, Button, K.J., Nijkamp P., and Pepping, G.J. (1997), *Meta-analysis in Environmental Economics*, Dordrecht: Kluwer Academic Publishers.
Blamey, R.K. and Common, M.S. (1999), 'Valuation and ethics in environmental economics', in J.C.J.M. van den Bergh (ed.), (1999), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Boulding, K.E. (1978), *Ecodynamics: A New Theory of Societal Evolution*, Beverly Hills: Sage Publications.
Boulding, K.E. (1981), *Evolutionary Economics*, Beverly Hills: Sage Publications.
Bromley, D.W. (1991), *Environment and Economy: Property Rights and Public Policy*, Oxford: Basil Blackwell.
Brouwer, R. (2000), 'Environmental value transfer: state of the art and future prospects', *Ecological Economics*, **32** (1), 137–52.
Buenstorf, G. (2000), 'Self-organization and sustainability: energetics of evolution and implications for ecological economics', *Ecological Economics*, **33**, 119–34.
Campbell, N.A. (1996), *Biology*, 4th edn, Melno Park, CA: Cummings Publishing.
Carson, R.T., Hanemann, W.M., Kopp, R.J., Presser, S. and Ruud, P. (1992), 'A contingent valuation study of lost passive use values resulting from the Exxon Valdez Oil Spill', Attorney General of the State of Alaska, Anchorage.
Clark, N., Perez-Trejo, F. and Allen, P. (1995), *Evolutionary Dynamics and Sustainable Development: A Systems Approach*, Aldershot, UK and Brookfield, USA: Edward Elgar.

Common, M. (1988), 'Poverty and progress revisited', in D. Collard, D.W. Pearce, and D. Ulph (eds), *Economics, Growth and Sustainable Environments*, New York: St Martin's Press, pp. 15–39.
Common, M. and Perrings, C. (1992), 'Towards an ecological economics of sustainability', *Ecological Economics*, **6**, 7–34.
Conlisk, J. (1989), 'An aggregate model of technical change', *Quarterly Journal of Economics*, **104**, 787–821.
Dawkins, R. (1976), *The Selfish Gene*, New York: Oxford University Press.
Day, R.H. and Walter, J.-L. (1989), 'Economic growth in the very long run: On the multiple-phase interaction of population, technology, and social infrastructure', in W.A. Barnett, J. Geweke and K. Shell (eds), *Economic Complexity, Chaos, Sunspots, Bubbles, and Nonlinearity*, Cambridge: Cambridge University Press, pp. 253–89.
Demeritt, D. and Rothman, D. (1999), 'Figuring the costs of climate change: an assessment and critique', *Environment and Planning A*, **31**, 389–408.
Diamond, J. (1997), *Guns, Germs and Steel: The Fates of Human Societies*, New York: W.W. Norton.
Durham, W.H. (1991), *Coevolution: Genes, Culture and Human Diversity*, Stanford, CA: Stanford University Press.
Ehrlich, P.R. and Raven, P.H. (1964), 'Butterflies and plants: a study in coevolution', *Evolution*, **18**, 568–608.
Ehrlich, P.R., Wolff, G., Daily, G.C., Hughes, J.B., Daily, S., Dalton, M. and Goulder, L. (1999), 'Knowledge and the environment', *Ecological Economics*, **30**, 267–84.
Epstein, C. and Axtell, R. (1996), *Growing Artificial Societies: Social Science from the Bottom Up*, Boston, MA: MIT Press.
Faber, M. and Proops, J.L.R. (1990), *Evolution, Time, Production and the Environment*, Heidelberg: Springer-Verlag.
Freeman III, A.M. (1993), *The Measurement of Environmental and Resource Values: Theory and Methods*, Washington, DC: Resources for the Future.
Georgescu-Roegen, N. (1971), *The Entropy Law and the Economic Process*, Cambridge, MA: Harvard University Press.
Georgescu-Roegen, N. (1972), 'Energy and economic myths', in N. Georgescu-Roegen (ed.) (1976), *Energy and Economic Myths*, New York: Pergamon.
Gould, S.J. (1988), 'On Replacing the Idea of Progress with an Operational Notion of Directionality', in M. Nitecki (ed.), *Evolutionary Progress*, Chicago: University of Chicago Press.
Gould, S.J. and Lewontin, R.C. (1979), 'The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme', *Proceedings of the Royal Society of London B*, **205**, 581–98.
Gowdy, J. (1994), *Coevolutionary Economics: The Economy, Society and the Environment*, Dordrecht: Kluwer Academic Publishers.
Gowdy, J.M. (1997), 'The Value of Biodiversity: Markets, Society and Ecosystems', *Land Economics*, 73 (1), 25–41.
Gowdy, J. (1999), 'Evolution, environment and economics', in J.C.J.M. van den Bergh (ed.), *Handbook of Environmental and Resource Economics*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Gowdy, J.M., and Ferrer-i-Carbonell, A. (1999), 'Toward consilience between biology and economics: the contribution of Ecological Economics' *Ecological Economics*, **29**, 337–48.
Hannan, M.T. and Freeman, J. (1989), *Organizational Ecology*, Cambridge, MA: Harvard University Press.

Hardin, G. (1968), 'The tragedy of the commons', *Science*, **162**, 1243–48.
Hausman, J.A. (ed.) (1993), *Contingent Valuation: A Critical Assessment*, Amsterdam: North-Holland.
Hinterberger, F. (1994), 'Biological, cultural and economic evolution and the economy–ecology relationship' in J.C.J.M. van den Bergh and J. van der Straaten (eds), *Toward Sustainable Development: Concepts, Methods and Policy*, Washington, DC: Island Press, ch. 3.
Holland, J.H. (1999), *Emergence: From Chaos to Order*, Boulder, Co: Perseus Book Group.
Holling, C.S. (1973), 'Resilience and stability of ecological systems', *Annual Review of Ecological Systems*, **4**, 1–24.
Holling, C.S. (1986), 'The resilience of terrestrial ecosystems: Local surprise and global change', in W.C. Clark and R.E. Munn (eds), *Sustainable Development of the biosphere*, Cambridge: Cambridge University Press.
Horgan, J. (1996), *The End of Science*, Reading, MA: Addison-Wesley.
Iwai, K. (1984), 'Schumpeterian dynamics, part I: an evolutionary model of innovation and imitation', *Journal of Economic Behaviour and Organization*, **5** (2), 159–90.
Jackson, T. (1999), 'Sustainability and the struggle for existence', Working paper, Centre for Environmental Strategy, University of Surrey, Guildford, UK.
Jackson, T. (2000), 'Why is ecological economics not an evolutionary science?', paper presented at the 3rd Biennial Conference of the European Society of Ecological Economics, Vienna, May.
Jackson, T. (2002), 'Evolutionary psychology in ecological economics: consilience, consumption and contentment', *Ecological Economics*, **41**, 289–303.
Janssen, M. (1998), *Modelling Global Change: The Art of Integrated Assessment Modelling*, Cheltenham, UK and Lyme, USA: Edward Elgar.
Janssen, M. and Vries, B. de (1998), 'The battle of perspectives: a multi-agent model with adaptive responses to climate change', *Ecological Economics*, **26**, 43–66.
Kahneman, D. and Tversky, A. (1979), 'Prospect theory: an analysis of decision under risk', *Econometrica*, **47**, 263–91.
Kauffman, S.A. (1993), *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford: Oxford University Press.
Kemp, R. (1997), *Environmental Policy and Technical Change: A Comparison of the Technological Impact of Policy Instruments*, Cheltenham, UK and Lyme, USA: Edward Elgar.
Kohlstad, C.D. (1994), 'The timing of CO_2 control in the face of uncertainty and learning', in E.C. van Ierland (ed.), *International Environmental Economics*, Amsterdam: Elsevier Science Publishers.
Ledyard, John O. (1995), 'Public goods: A survey of experimental research', in John H. Kagel and Alvin E. Roth (eds), *The Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press, pp. 111–94.
Levin, S.A. et al. (17 authors) (1998), 'Resilience in natural and socioeconomic systems', *Environment and Development Economics*, **3** (2), 222–35.
Lumsden, C.J. and Wilson, E.O. (1981), *Genes, Mind and Culture: The Coevolutionary Process*, Cambridge, MA: Harvard University Press.
Martin, F. (1989/1992), 'Common-pool resources and collective action: a bibliography', Vols 1 and 2, Workshop in Political Theory and Policy Analysis, Bloomington, IN: Indiana University.
Maynard Smith, J. and Szathmáry, E. (1995), *The Major Transitions in Evolution*, Oxford: Oxford University Press.

McNeill, J.R. (2000), *Something New Under the Sun: An Environmental History of the Twentieth-Century World*, New York: W.W. Norton & Co.
Mokyr, J. (1990a), 'Punctuated equilibria and technological progress', *American Economic Review, Papers and Proceedings*, **80** (2), 350–54.
Mokyr, J. (1990b), *The Lever of the Riches: Technological Creativity and Economic Progress*, Oxford: Oxford University Press.
Mulder, P. and Bergh, J.C.J.M. van den (2001), 'Evolutionary economic theories of sustainable development', *Growth and Change*, **32** (4), 110–34.
Munro, A. (1997), 'Economics and biological evolution', *Environmental and Resource Economics*, **9**, 429–49
Nelson, R. and Winter, S. (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
Noailly, J., Bergh, J.C.J.M. van den and Withagen, C. (2003a), 'Evolution of harvesting strategies: replicator and resource dynamics', *Journal of Evolutionary Economics*, **13** (2), 183–200.
Noailly, J., Bergh, J.C.J.M. van den, and Withagen, C. (2003b), 'Spatial evolution of social norms in a common pool resource game', unpublished paper.
Nordhaus, W.D. (1994), *Managing the Global Commons: The Economics of Climate Change*, Cambridge, MA: MIT Press.
Norgaard, R.B. (1984), 'Coevolutionary development potential', *Land Economics*, **60**, 160–73.
Norgaard, R.B. (1994), *Development Betrayed: The End of Progress and a Coevolutionary Revisioning of the Future*, London and New York: Routledge.
Norton, B., Costanza, R. and Bishop, R.C. (1998), 'The evolution of preferences. Why "sovereign" preferences may not lead to sustainable policies and what to do about it', *Ecological Economics*, **24**, 193–211.
O'Connor, M. (1991), 'Entropy, structure, and organisational change', *Ecological Economics*, **3**, 95–122.
Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action New York*. Cambridge: Cambridge University Press.
Perrings, C. (1998), 'Resilience in the dynamics of economy–environment systems', *Environmental and Resource Economics*, **11**, 503–20.
Perrings, C., Williamson, M.H. and Dalmazzone, S. (eds) (2001), *The Economics of Biological Invasions*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Pimm, S.L. (1984), 'The complexity and stability of ecosystems', *Nature*, **307**, 321–6.
Potts, J. (2000), *The New Evolutionary Microeconomics: Complexity, Competence, and Adaptive Behavior*, Cheltenham, UK and Northampton, USA: Edward Elgar.
Roe, E.M. (1996), 'Sustainable development and Girardian economics', *Ecological Economics*, **16**, 87–93.
Ruse, M. (1999), *Mystery of Mysteries: Is Evolution a Social Construction?* Cambridge, MA: Harvard University Press.
Schneider, E.D. and. Kay, J.J (1994), 'Life as a manifestation of the second law of thermodynamics', *Mathematical and Computer Modelling*, **19** (6–8), 25–48.
Sen, A. (1993), 'On the Darwinian view of progress', *Population and Development Review*, **19** (1), 123–37.
Sethi, R. and Somanathan, E. (1996), 'The evolution of social norms in common property resource use', *American Economic Review*, **86** (4), 766–88.
Silverberg, G. and Verspagen, B. (1994), 'Collective learning, innovation and

growth in a boundedly rational, evolutionary world', *Journal of Evolutionary Economics*, **4**, 207–26.

Silverberg, G., Dosi, G. and Orsenigo, L. (1988), 'Innovation, diversity and diffusion: a self-organization model', *Economic Journal*, **98**, 1032–54.

Somit, A. and Peterson, S. (1989), *The Dynamics of Evolution: The Punctuated Equilibrium Debate in the Natural and Social Sciences*, Ithaca, NY: Cornell University Press.

Spash, C.L. and Hanley, N. (1995), 'Preferences, information and biodiversity preservation', *Ecological Economics*, **12**, 191–208.

Stern, D. (1997), 'Limits to substitution and irreversibility in production and consumption: a neoclassical interpretation of ecological economics', *Ecological Economics*, **22**, 197–215.

Strickberger, M.W. (1996), *Evolution*, 2nd edn, Sudbury, MA: Jones and Bartlett Publishers.

Toman, M.A., Pezzey, J. and Krautkraemer, J. (1995), 'Neoclassical economic growth theory and "sustainability",' in D. Bromley (ed.), *Handbook of Environmental Economics*, Oxford: Blackwell.

van der Voet, E., Huele, R. and Stevers, R. (2000), 'Evolutionary predators, prey, pets and pests: evolutionary reactions on human dominance in the biosphere', unpublished, Centre for Environment Science, University of Leiden, The Netherlands.

Wilkinson, R. (1973), *Poverty and Progress: An Ecological Model of Economic Development*, London: Methuen and Co.

Wilson, E.O. (1975), *Sociobiology: The New Synthesis*, Cambridge, MA: Harvard University Press.

Wilson, E.O. (1998), *Consilience*, New York: Alfred Knopf.

Index